PROCEEDINGS
OF THE
NATIONAL ACADEMY OF SCIENCES
Volume 23 June 15, 1937 Number 6

CAROTENOIDS AND OTHER LIPOID-SOLUBLE PIGMENTS IN THE SEA AND IN DEEP MARINE MUD

By Denis L. Fox

Scripps Institution of Oceanography of the University of California, La Jolla, California

Communicated May 4, 1937

Marine animals of the euphotic zone derive their carotenoids by consuming the plants found there, or by eating other animals which so derive their food. On the other hand, the origin of the relatively rich stores of carotenoids possessed by certain benthonic animals requires special consideration, since animals, unlike plants, probably cannot synthesize carotenoids *de novo* but must obtain such pigments from an ultimate plant source.

Murray and Hjort (1912) discuss at some length the fact that the darkest colored marine animals commonly occur at the greater depths. The question of the food sources supplying carotenoids to the numerous yellow, orange and red colored benthonic animals, presents some interesting problems relating to nutrition.

These and kindred questions have led to an interest in the possible kinds and relative quantities of carotenoid pigments in microscopic organisms along with fecal and other detrital material suspended in ocean waters, or incorporated in marine bottom muds.

Carotenoids have been reported in certain soils (Baudisch and von Euler, 1935) and have been mentioned as occurring in marine sediments (Trask, 1932). It has long been known that carotenoids occur in kelps and microscopic marine algae, as in all other photosynthetic plants, also in numerous bacteria, fungi and planktonic animals (Palmer, 1922; Chargaff, 1933; Karrer and Solmssen, 1936).

In this laboratory, carotenoids have been demonstrated in material filtered from suspension in sea water, in natural marine sediments and in sand which had been stored for some weeks under running sea water, where it accumulated much organic detritus. Carotenoids, accompanied sometimes by chlorophyll, have also been demonstrated here in the feces of several kinds of marine animals, such as the kelp-eating crab, *Epialtus;*

phytoplankton-feeding mollusks such as the mussel, *Mytilus;* certain bottom-feeding or ilytrophic[1] animals such as the sea cucumber, *Stichopus;* and carnivorous surf perches (Embiotocidae) which devour small shrimps rich in carotenoids (Young and Fox, 1936). The greenfish, *Girella nigricans*, the Pacific killifish, *Fundulus parvipinnis* and the brilliantly orange-colored Garibaldi, *Hypsypops rubicundus*, may show copious plant remains in the intestine and feces. (Sumner and Fox, 1935a, 1935b).

Drummond and Macwalter (1935) found that the feces of Antarctic whales contain large amounts of the red acidic carotenoid astacene, derived from the enormous numbers of pelagic copepods consumed by these marine mammals.

General Procedure.—In all the work the general procedure commonly used for studies of carotenoids was employed. This entailed (1) a preliminary extraction of the material with 95% ethyl alcohol, followed if necessary by additional extraction with ligroin (Eastman Kodak Co., P. 513, boiling range 70–90°C); (2) quantitative separation of the extracted pigments into ligroin-soluble and (90%) methyl alcohol-soluble fractions; (3) hydrolysis of xanthophyll esters in the ligroin-soluble fraction by treatment with hot alcoholic KOH under nitrogen gas; (4) chromatographic separation of (*a*) the carotenes by passing their dry, alcohol-free ligroin solutions through Tswett columns of solid $Ca(OH)_2$ (Miller, 1935), and (*b*) the xanthophylls by passing their dry, alcohol-free ligroin solutions through Tswett columns of solid $CaCO_3$ (Strain, 1936; Zechmeister, 1934, and others); finally the separated pigments from each group were transferred to pure CS_2 (Merck's) and their absorption spectra were determined with a Bausch and Lomb spectrophotometer.

Sea Water.—Sea water pumped from the ocean to a settling tank, and issuing thence from an aquarium tap was passed through a column of cotton wads packed into a glass tube. Approximately 4000 liters of sea water were thus filtered, and the wads bearing the filtered material were extracted of carotenoids in the regular manner (Fox, *et al.*, 1936). The combined yield was about 0.123 mg. carotenoids (calculated from our xanthophyll curve (S. and F., 1935b), of which approximately 18.5% was carotene-like in nature, not being adsorbed by a dry Tswett column of $CaCO_3$. A part of the "carotene" fraction may have been composed of xanthophyll esters, since these behave like carotenes both in the partition test between ligroin and 90% CH_3OH, and in not being adsorbed by $CaCO_3$. Furthermore, xanthophyll esters have been found to be abundant in the carotenoids of marine animals (Sumner and Fox; Young and Fox; *op cit.*, Fox, 1936) and are present also in marine mud. The quantity of carotenoids in sea water will, needless to say, vary greatly with the numbers of planktonic organisms and with the amounts of suspended organic detritus, and no single determination can be taken as representative.

The question of carotenoids present in suspended detritus and planktonic organisms in the sea was incidentally approached by an indirect method, namely, by analyzing the feces of the plankton-feeding mussel, *Mytilus californianus* (see Fox, *et al.*, 1936). The particular lot of feces analyzed yielded no chlorophyll, approximately 0.050% xanthophylls (dry feces) and only some 0.003% "carotenes."

Extracts of the "liver" or digestive diverticulum of the mussel yielded relatively large quantities of carotenoid pigments, of which by far the greater amount were of the xanthophyll series.

At the time that the analyses were made of suspended matter in sea water and of mussel feces, no spectroscopic equipment was available, hence the different carotenoids were not identified. Several carotenoids were doubtless present, however, since the xanthophylls adsorbed by the $CaCO_3$ column appeared as different illy defined bands of rose, salmon and pinkish colors. This would be expected, since the source of the carotenoids was in the remains of photosynthetic plants. Furthermore, the marine bottom showed several different carotenoids (see below).

Marine Mud.—Mud cores were obtained aboard our late vessel, the "Scripps," at various points in the ocean off San Diego, including one station over a deep trough on the continental shelf, located at 32° 37.6′ N, 118°09′ W some 46 miles west of Point Loma. The sample studied was secured on January 26, 1936, at a depth of 2007 meters (approximately 1100 fathoms). Oceanographic data show that oxygen and temperature figures at 2000 meters would differ only slightly from those at 1900 meters (conc. O_2 = 1.4 cc./liter; T = 2.64°C). Sections of the core were placed in dark bottles, and stored next day in a refrigerator at temperatures at or below 0°C until ready to be analyzed.

Part of a core from the top four inches was taken for preliminary investigation. The mud was extremely fine and sticky, of a dark greenish brown color and showed a water content of 63.2%. Repeated extraction of 47.4 grams of undried mud with successive additions of absolute ethyl alchol removed all of the pigment and gave a yellow solution. No chlorophyll was detected in any of the extract.

The epiphasic fraction (soluble in ligroin, and referred to as the "carotene" fraction) showed 0.025 mg. of pigment calculated as carotene per 100 grams of undried mud, by the Ives tint photometer (see Sumner and Fox, 1933). The hypophasic fraction (soluble in 90% CH_3OH, and consisting of xanthophylls) showed 0.047 mg. of xanthophyll per 100 grams of undried mud, when analyzed by the same method. The "carotene" fraction contained some xanthophyllic compounds, since, although it was entirely epiphasic, a trace of pigment was adsorbed on a Tswett column of $CaCO_3$. All, or all excepting the faintest trace of pigment, was absorbed on $CaCO_3$ from the originally hypophasic fraction (xanthophylls).

Another section of the same core, consisting of mud from the 8-inch to the 16-inch depth, was subjected to a somewhat more searching qualitative analysis for carotenoids; 197.5 grams of this mud, which was of extreme fineness and of the same general appearance as the first sample, had been stored for 11 months in the dark in a refrigerator at temperatures close to 0°C and below so that latterly ice crystals had formed throughout the masses. It was considered that under these conditions, little or no oxidative destruction of the carotenoids would be likely to have occurred.

The main CH_3OH extract was a golden-yellow color. This extract was resolved in the usual manner into a ligroin-soluble fraction and a fraction selectively soluble in 90% CH_3OH. The xanthophyll-like pigments (soluble in 90% CH_3OH) were than transferred to ligroin by diluting the alcohol in the system and extracting quantitatively with successive ligroin washes. The fractions will be considered in order.

Ligroin-Soluble Fraction.—This fraction contained, along with a number of carotenoids, a dirty greenish component which was readily adsorbed by $CaCO_3$, remained entirely epiphasic and gave a greenish yellow solution in ligroin. Evaporation of the solvent left dark brown, kelp-colored oily drops. The material was highly resistant to long treatment with alcoholic KOH; it gave a yellow-orange solution in CS_2, but exhibited no absorption maxima on spectrophotometric analysis in the green to violet region, giving only an increasing absorption toward the shorter wave-lengths. This pigment was neither a carotenoid nor chlorophyll, but may have been some derivative of one of these.

An epiphasic carotenoid which resisted hydrolysis and was unadsorbed by either $CaCO_3$ or $Ca(OH)_2$ gave an orange solution in CS_2 and showed absorption maxima of the average values: *I*, 507.5 mμ; *II*, 477.5 mμ. The absorption maxima of α-carotene are given by Smith (1936) as *I*, 507.4 mμ; *II*, 476.6 mμ. This close agreement, together with similarities in chemical behavior, indicated this pigment to be α-carotene separated from β-carotene by adsorption of the latter upon $Ca(OH)_2$. The other epiphasic carotenoid which resisted hydrolysis and was very poorly adsorbed by $CaCO_3$ gave a blood-red solution in CS_2 and possessed the following absorption bands approximately: *I*, 512.5–515.0 mμ; *II*, 485.0 mμ; these absorption maxima agree rather closely with those given by Smith for β-carotene: *I*, 512.5 mμ; *II*, 483.4 mμ. The carotene-like chemical behavior and the close agreement in the adsorption spectrum suggest that the fraction consisted of β-carotene.

Hydrolysis of the components containing the α- and β-carotenes during their purification yielded small quantities of xanthophyll-like pigments which were adsorbed by $CaCO_3$ or were preferentially soluble in 90% CH_3OH in the partition operation. These components were present in too small amounts for spectroscopic analysis.

The ligroin soluble fraction of the extract of marine mud contained: an unidentified greenish brown pigment; α-carotene, β-carotene, small quantities of xanthophyll esters and possibly traces of partially oxidized carotenes.

Alcohol-Soluble Fraction.—Some green-brownish material was recovered from the chromatographic separation of the pigments of the hypophasic component. The material was readily adsorbed by $CaCO_3$, nearly insoluble in ligroin, but readily so in CH_3OH at room temperature. It gave a greenish brown solution in chloroform; evaporation of the solvent yielded an oily, sepia-brown residue which was not completely dissolved by diethyl ether or by CS_2. The chloroform solution gave a pale fluorescent appearance when illuminated by a narrow beam of light. Examined in the spectrophotometer, this solution showed no maxima in the region from 435.0 mμ up to 540.0 mμ, but only a gradually increasing absorption toward the violet. It gave a slight maximal absorption down in the red at about 670.0 mμ, which is close to the one of the chief maxima for chlorophyll. Its chemical properties and color were such that it could not have been chlorophyll. It may have been allied to or derived from chlorophyll.

Two carotenoids were recovered from the fraction; one which was relatively poorly adsorbed on $CaCO_3$, advancing through the column as a yellow band, gave a yellow-orange solution in CS_2, and showed indications of absorption maxima at about *I*, 505 mμ; *II*, 480 mμ; the other component was firmly adsorbed to $CaCO_3$ as a red band, was eluted in the regular manner, gave a yellow-orange solution in CS_2 and indicated absorption maxima at approximately *I*, 502.5–505 mμ; *II*, 475–480 mμ.

Earlier spectrophotometric examination of spinach xanthophylls (from Eimer and Amend) showed maxima at *I*, 505 mμ; *II*, 475 mμ (unpublished experiments). Chromatographic separation of the commercial spinach xanthophylls gave other components whose maxima lay at slightly shorter wave-lengths, viz., *I*, 500 mμ; *II*, 470 mμ (corresponding closely to violaxanthin in absorption maxima, in ready adsorption by $CaCO_3$, and in giving a blue color in ether with strong HCl); *I*, 505 mμ; *II*, 475 mμ, corresponding rather closely to the adsorption bands of lutein (Kuhn and Smakula, 1931) and very probably identical with this xanthophyll, the chief xanthophyll of spinach. Other absorption maxima lay at points intermediate between the above values, and were probably indicative of mixtures and possible isomers which were unidentified.

Strain (1936) points out that there are at least twelve and probably more xanthophylls in green leaves, and that these are separable by various refined techniques, including chromatographic adsorption. It is possible that a similar plurality of xanthophylls would be encountered in marine plant material as well.

The rather close agreement between the properties of the xanthophylls

from the marine mud and those of the chief xanthophyll of spinach would suggest that we were dealing with lutein, and possibly with closely similar isomers, such as isolutein (Strain, 1936) as well.

Summary and Conclusions.—In the marine euphotic zone a plentiful supply of carotenoids is assured to animals through the ultimate agency of plants. At great depths, carotenoids are available in the diet of mud eaters and animals which prey upon them.

Carotenoids are present in material filtered from sea water, including microplankton, fecal matter and other organic detritus. Marine muds from depths of 1000 fathoms also contain carotenoids which were demonstrated in cores from the mud surface to a depth of 16 inches.

Unesterified xanthophylls seemingly preponderate over carotenes and esterified xanthophylls in carotenoids from both water and mud. Alpha- and beta-carotene, lutein and other xanthophyllic carotenoids both free and esterified were isolated from the marine mud. Chlorophyll was not detected in the mud, but greenish brown non-carotenoid pigments were encountered which may have been related to chlorophyll.

The supplying of carotenoids to marine muds by the settling of dead organisms, feces and other detritus from the water above is doubtless a continual process. Carotenoid pigments are perhaps synthesized also by endemic achlorophyllic microörganisms. It would seem likely, *a priori*, that carotenoids found embedded a foot or more beneath the surface of deep marine muds must be of very long standing.[2]

[1] The term *ilytrophism* and co-derivatives from the same Greek roots, proposed to signify the mud-eating habit, are defined in *Science*, **84**, 437 (1936).

[2] Since this paper was submitted, other work by P. D. Trask has come to the attention of the writer, which seems to lend support to this conclusion regarding age; see "Inferences about the Origin of Oil as Indicated by the Composition of the Organic Constituents of Sediments," U. S. G. S. Prof. Paper No. 186H. U. S. Govt. Printing Office, Washington, D. C., 1937.

Baudisch, O., and H. v. Euler, *Archiv. Kemi. Mineral. Geol.*, **11A**, No. 21 (1935).

Chargaff, E., *Compt. rend.*, **197**, 946–948 (1933).

Drummond, J. C., and R. J. Macwalter, *Jour. Exptl. Biol.*, **12**, 105–107 (1935).

Fox, D. L., *Proc. Nat. Acad. Sci.*, **22**, 50–54 (1936).

Fox, D. L., *et al.*, *Bull. Scripps Inst. of Oceanog. of Univ. of Calif., Tech. Series*, **4** 1–64 (1936).

Karrer, P., and U. Solmssen, *Helv. Chim. Acta*, **19**, 3–5 (1936).

Kuhn, R., and A. Smakula, *Zeit. physiol. Chem.*, **197**, 161 (cited by Zechmeister) (1931)

Miller, E. S., *Bot. Gaz.*, **96**, 447–467 (1935).

Murray, Sir J., and J. Hjort, *The Depths of the Ocean.* Macmillan and Co. Ltd., London (1912).

Palmer, L. S., *Carotinoids and Related Pigments*, A. C. S. Monograph Series. Chemical Catalog Company, N. Y. (1922).

Smith, James H. C., *Jour. Amer. Chem. Soc.* **58**, 247–255 (1936).

Strain, H. H., *Science*, **83**, 241–242 (1936).

Sumner, F. B., and D. L. Fox, *Jour. Exptl. Zoöl.*, **66**, 263-301 (1933).
Sumner, F. B., and D. L. Fox, *Proc Nat. Acad. Sci.*, **21**, 330-340 (1935).
Sumner, F. B., and D. L. Fox. *Jour. Exptl. Zoöl.*, **71**, 101-123 (1935b).
Trask, P. D., *et al.*, *Origin and Environment of Source Sediments of Petroleum*. Gulf Pub. Co., Houston, Texas (1932).
Young, R. T., and D. L. Fox, *Biol. Bull.*, **71**, 217-237 (1936).
Zechmeister, L., *Carotinoide*, J. Springer, Berlin (1934).

INTERNUCLEAR DISTANCE IN OXYGEN MOLECULES

BY HAROLD D. BABCOCK

MOUNT WILSON OBSERVATORY, CARNEGIE INSTITUTION OF WASHINGTON

Read before the Academy Tuesday April 27, 1937

In his report on Band Spectra of Diatomic Molecules[1] Jevons remarks "the difference of moment of inertia of two isotopic molecules arises almost entirely from a difference of 'reduced' or 'effective mass'—since the other factor, internuclear distance—depends upon the extranuclear electric field, which must be very nearly, though not quite, the same for both molecules." Data adequate for a rigorous test of the equality of the extranuclear fields were apparently not available, since he did not amplify the statement quoted.

Measurements of twelve absorption bands of atmospheric oxygen have now been completed which permit some explicit qualification of Jevons' statement. My observations deal with eight bands associated with the ordinary molecule, $O^{16}O^{16}$, and two each from the isotopic molecules $O^{16}O^{17}$ and $O^{16}O^{18}$. The electronic transitions are $^3\Sigma \rightarrow {}^1\Sigma$ for eleven of the bands in the spectral region λλ 5800–7800. The twelfth band, near λ 12680, corresponds to the electronic transition $^3\Sigma \rightarrow {}^1\Delta$, which is observable only for $O^{16}O^{16}$. Altogether, about 575 oxygen lines have been measured and assigned to their places in these twelve bands, but less than one-half have been selected for the purposes of this paper. The range of intensities is so great that air-paths as short as 30 meters have been used for the strongest lines, and as long as can be obtained on Mount Wilson for the weakest. Effective length of air-path at sunset cannot be accurately computed because of uncertainty in the effect of refraction, but the order of magnitude is probably about 200 km.

The analysis of the measurements has been based on Mulliken's interpretation[2] of $^3\Sigma \rightarrow {}^1\Sigma$ in O_2, extended as necessary to take account of differences between the ordinary symmetrical molecule and the unsymmetrical isotopic molecules. By well-known methods some of the constants of the

molecules are derived from the positions of the individual lines in each band, with precaution to avoid the use of blended lines or those associated with the higher states of rotation.

For the purposes of this paper the most interesting constant obtained directly from the measurements is B_v, the coefficient of the principal term in the expression for the angular momentum of the molecule, in the vibrational state corresponding to v, as it rotates about its center of mass. The effects of a slight departure from rigidity are removed analytically and values of the moment of inertia and the internuclear distance, r_v, are derived from B_v by the usual simple relations. Results obtained in this way are in good agreement with those found by Dieke and Babcock,[3] but the uncertainty is now appreciably less.

TABLE I

INTERNUCLEAR DISTANCE IN OXYGEN MOLECULES, $r_v \times 10^8$ cm.

ELECTRONIC STATE	VIB. STATE	E. P. VOLTS	$O^{16}O^{16}$	$O^{16}O^{17}$	$O^{16}O^{18}$
$^3\Sigma$	0	0	1.20756	1.2077	1.20754
			±0.00001	±0.0003	±0.00004
	1		1.21425		
			±0.00004		
$^1\Delta$	0	0.97	1.2167		
			±0.0002		
$^1\Sigma$	0	1.62	1.22758	1.2274	1.2273
			±0.00002	±0.0002	±0.0002
	1		1.23579	1.2356	1.2351
			±0.00004	±0.0003	±0.0001
	2		1.24409		
			±0.00002		
	3		1.25284		
			±0.00001		

Table 1 shows the values of r_v associated with various electronic and vibrational states of each type of oxygen molecule. Several of the results given here are derived from two or three bands yielding practically coincident values. The observed dimensions of $O^{16}O^{16}$ and $O^{16}O^{18}$ in the $^3\Sigma$ state are seen to be the same within about 1 part in 50,000, but a decrease of internuclear distance, r_v, for the heavier molecules is plainly shown in the $^1\Sigma$ state, and the internal probable errors, derived from least-squares reduction of the observations, confirm its reality. The mean value of the change from $O^{16}O^{16}$ to $O^{16}O^{18}$ for the two cases shown here is 4 parts in 10,000, with a probable error of about 25 per cent.

Table 1 also shows that the molecule $O^{16}O^{16}$, when compared in homologous vibrational states, increases in size approximately in proportion to the $^3/_2$ power of its change of electronic excitation. Furthermore, the

increase of size with augmented vibrational energy is less for the normal state, $^3\Sigma$, than for the excited state, $^1\Sigma$, and in the latter case is proportional to the vibrational quantum number.

For $O^{16}O^{16}$ the data are sufficient for an accurate determination of an auxiliary constant, usually called α, permitting a further slight reduction of the internuclear distance to that value, r_e, which it would have in a hypothetical state entirely free from vibration, both for the normal molecule and for the excited state, $^1\Sigma$. These values of r_e are given in table 2. For $O^{16}O^{17}$ and $O^{16}O^{18}$ a similar though less complete treatment is possible only for the $^1\Sigma$ state, because the necessary bands have not yet been observed on account of their low intensity. The observations indicate that in the $^1\Sigma$ state r_e is the same for the isotopic molecules as for $O^{16}O^{16}$ to 1

TABLE 2

INTERNUCLEAR DISTANCE IN HYPOTHETICAL OXYGEN MOLECULES HAVING NO VIBRATIONAL ENERGY, $r_e \times 10^5$ cm.

ELECTRONIC STATE	$O^{16}O^{16}$	$O^{16}O^{17}$	$O^{16}O^{18}$
$^3\Sigma$	1.20426††	1.2044*	1.20439*
	±0.00002	±0.0003	±0.00007
		1.2043**	1.20407**
		±0.0002	±0.00005
$^1\Sigma$	1.22349††	1.2234†	1.2234†
	±0.00005	±0.0002	±0.0001
		1.2233**	1.2230**
		±0.0002	±0.0001

* α obtained indirectly by proportion.
** α assumed same as for $O^{16}O^{16}$.
† α observed from meager data.
†† α accurately determined.

part in 12,000. But the magnitude of the auxiliary constant, α, is apparently slightly different for each kind of molecule and some assumption becomes necessary before r_e can be derived for the heavier molecules in the $^3\Sigma$ state. If for each of these molecules α is supposed to have the same ratio to its value for $O^{16}O^{16}$ in the ground state that it has in the $^1\Sigma$ state, the dimensions are found to be the same to 1 part in 12,000, as shown in table 2. If, however, we take α for the heavy molecules the same as for $O^{16}O^{16}$, ignoring the few observations which indicate small variations, the internuclear distance, r_e, shows variations similar to those already seen in the actual molecules, listed in table 1.

In these calculations the masses found spectroscopically by Birge and Babcock[4] were used, namely 17.008 and 18.0065. It is well known that observations of nuclear disintegration have led to masses considerably

lower than those used here for the isotopes of oxygen. If the disintegration value 18.0037 is substituted for 18.0065, however, the result is an increase of only 1 part in 24,000 in r_e. It thus seems probable that errors in the masses of O^{17} and O^{18} do not seriously affect the results given here.

It should be noted that the calculations of internuclear distance described here have been made by means of relations known to be approximate, though presumably rigorous enough for most previous applications. It is possible that an extension of the theory of molecular structure as derived from band spectra might now be susceptible of quantitative test, at least for the case of oxygen. Special interest would be found in the analysis of refined measurements on molecular hydrogen and deuterium.

Full details are being prepared for publication elsewhere.

[1] Physical Society, Cambridge (1932).
[2] *Phys. Rev.*, **32**, 880–887 (1928).
[3] These PROCEEDINGS, **13**, 670–678 (1927).
[4] Unpublished results and *Phys. Rev.*, **37**, 233 (1931).

BUD DEVELOPMENT IN LILIUM HARRISII FOLLOWING TREATMENT WITH INDOLEACETIC ACID

BY J. M. BEAL

DEPARTMENT OF BOTANY, THE UNIVERSITY OF CHICAGO

Communicated May 8, 1937

Previous reports dealing with the effects of growth promoting substances on higher plants have indicated the inhibition or marked retardation of lateral bud development when the terminal bud is excised and the substance is applied to the cut surface of the stem. This preliminary report dealing with the application of indoleacetic acid on certain species of Lilium shows that lateral bud development may be stimulated by this treatment. Whether these buds are adventitious or not is yet to be determined.

Experiments on the Red Kidney variety of the bean (Kraus, Brown and Hamner[1]) suggested the advisability of testing the indoleacetic acid on some lilies then in active growth in the greenhouses, and the same mixture that they used (30 mg. of indoleacetic acid to 1 gram of lanolin) was applied to *Lilium philippinense formosanum* and *L. Harrisii*. Some of these plants showed immature floral buds, while others were strictly vegetative. Both kinds of stems were used. When the stem was terminated by a single floral bud it was cut off about 1 cm. below the receptacle, while if more than one bud was present the stem was cut just below the lowermost pedicel. In

the strictly vegetative stem, since the terminal bud consists of a much shortened axis surrounded by a number of leaves of varying sizes, it was necessary to cut the stem a few millimeters back of the growing point. The lanolin mixture was then spread over the entire cut surface in both types of stems. The responses were essentially identical in both, but markedly different in the two species used. In both species there was little immediate response in so far as enlargement of the stem tip was concerned, although somewhat more rapid in *L. Harrisii* than in *L. philippinense formosanum*. In the latter, small protuberances began to appear on the sur-

FIGURE 1 FIGURE 2

Figures 1–2. Photographs (× 2/3) of two plants of *L. Harrisii*.

Figure 1. Treated plant which has produced four buds (bulbils), one in the axil of the uppermost leaf which was partly severed in cutting off the terminal bud, two in the axil of the second leaf and one in the axil of the third leaf. Figure 2. Check plant.

face of the treated stems in from ten days to three weeks after treatment. The region of response extended from approximately the surface of application to as far down as 2 to 2.5 cm. from it. Some of these stems have since been sectioned and the protuberances have been shown to be due to developing adventitious roots. No buds or bud primordia were visible in the leaf axils at the time of treatment and none appeared subsequently, although some of the treated plants were allowed to remain undisturbed for more than two months.

Fifteen plants of *L. Harrisii*, after excision of the terminal buds, were treated with the lanolin mixture and four were left as checks. Callus tis-

sue began to develop on the surface of application of some of the treated stems in a few days, and a much smaller number of roots developed than with *L. philippinense formosanum*. But by far the most striking development was the appearance of buds, not visible at the time of treatment, in the axils of the upper two or three leaves (Fig. 1), and in at least two plants buds have developed from callus tissue on the cut surface of the stems. In none of the four check plants have buds developed at or near the cut surface or in the leaf axils even after a period of more than 12 weeks (Fig. 2). In three of the treated plants, two buds have developed in the axil of the uppermost leaf and a total of four buds has been produced by each of these plants. A number of the treated stems were killed and fixed at various intervals following the treatment, but seven of the plants have remained undisturbed for further observations. These seven plants now show a total of 17 buds.

The buds may develop near the side of the leaf attachment to the stem as frequently as they do in the strictly axillary position. When two are present in the same leaf axil, one being strictly axillary and the other accessory in position, they may be equal in size or either may outgrow the other. Apparently the one which develops first will grow somewhat more rapidly than that formed later.

Two or three weeks after the buds become evident they begin to assume a bulbous form. At this time, and subsequently, they resemble closely the bulbils which form naturally near the bases of the stems of untreated plants of this species. This is clearly shown in figure 1 where the largest of the bulbils has some leaves nearly one and a half inches in length. Roots have also developed from several of the bulbils. Attempts are now being made to produce plants from some of the bulbils with the hope that they may furnish material for cytological studies. Critical studies of the histological and cytological changes induced by the treatment with the lanolin mixture on both species are now in progress, and a more detailed report will be presented at a later date.

[1] Kraus, Brown and Hamner, *Bot. Gaz.*, **98**, 370–420 (1936).

THE DISTRIBUTION OF GENE FREQUENCIES IN POPULATIONS

BY SEWALL WRIGHT

DEPARTMENT OF ZOÖLOGY, UNIVERSITY OF CHICAGO

Read before the Academy, April 26, 1937

The effects of the various evolutionary factors—mutation, cross breeding, selection and inbreeding—can be reduced to common terms by considering the rates of change which they tend to bring about in the relative frequencies of alleles.[1] In the absence of such factors, there is constancy of gene frequencies from the symmetry of the Mendelian mechanism.

The frequency (q) of a given gene changes at the rate $\Delta q = -uq$ per generation under recurrent mutation of the gene to alleles at the rate u. Mutation in the opposite direction at the average rate v per generation changes the gene frequency at the rate, $\Delta q = v(1 - q)$.

If a certain gene has the frequency q in a local population but q_t in the species as a whole, exchange of the proportion m of the local population with an equal number of random individuals from the whole species leads to change of gene frequencies in the former at the rate $\Delta q = -m(q - q_t)$. Cross breeding is, however, most likely to be with neighboring populations which differ but little in value of q. In this case the coefficient m is only a small fraction of the actual amount of exchange. There may be other complications such as selective immigration or emigration, but the above simple form will suffice here to illustrate cross breeding or migration pressure.

The simplest kind of selection is that in which the heterozygote is exactly half way between the two homozygotes in the extent per individual to which it contributes to the next generation. The selective value of zygotes (relative to a certain standard) will be designated w and the mean value for a population, $\bar{w}$.

ZYGOTE	FREQUENCY	w	
AA	$(1 - q)^2$	1	$\bar{w} = 1 - 2sq$
AA'	$2q(1 - q)$	$1 - s$	$\frac{d\bar{w}}{dq} = -2s$
$A'A'$	q^2	$1 - 2s$	

$$\Delta q = \frac{(1 - s)q(1 - q) + (1 - 2s)q^2}{1 - 2sq} - q =$$

$$\frac{-sq(1 - q)}{1 - 2sq} = \frac{q(1 - q)}{2\bar{w}} \frac{d\bar{w}}{dq}. \qquad (1)$$

In the more general case in which $\bar{w}$ is not related linearly to q, the momentary selective advantage $(-2s)$ of replacing A by A' is still given

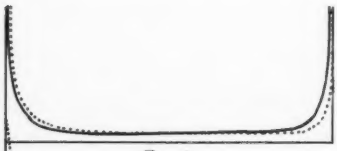

Fig. 1

Fig 4

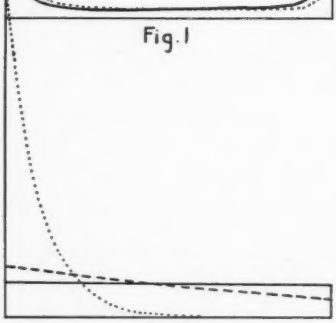

Fig. 2

Fig. 5

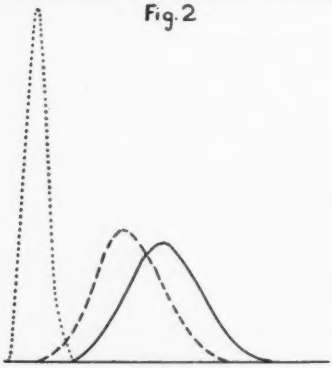

Fig. 3

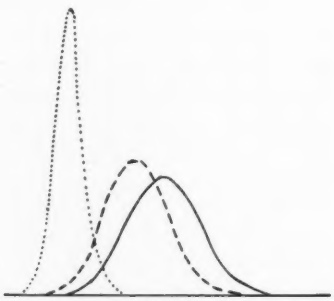

Fig. 6

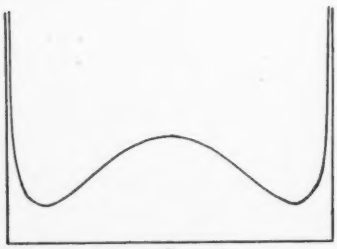

Fig. 7

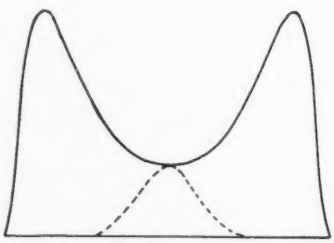

Fig. 8

(Captions for figures on opposite page.)

by $d\bar{w}/dq$ and the value of Δq is given by the same formula as above in terms of w, $d\bar{w}/dq$ and q.

ZYGOTE	FREQUENCY	w	
AA	$(1-q)^2$	1	$\bar{w} = 1 - 2s_1q(1-q) - s_2q^2$
AA'	$2q(1-q)$	$1-s_1$	$\frac{d\bar{w}}{dq} = -2s_1 + 2(2s_1 - s_2)q$
$A'A'$	q^2	$1-s_2$	

$$\Delta q = \frac{q(1-q)}{2\bar{w}}\frac{d\bar{w}}{dq}. \qquad (2)$$

Still more generally,[2] selective values depend on the interactions of the entire system of genes. It is the harmonious development of all characteristics that determines the success of an organism, not the absolute grades

---

CAPTIONS FOR FIGURES ON OPPOSITE PAGE

Figures 1 to 3. Some of the forms taken by the distribution of gene frequencies in the case of no dominance. $(\varphi(q) = Ce^{4Nsq}q^{4Nv-1}(1-q)^{4Nu-1})$. Mutation rates are assumed constant and equal $(u = v)$. Effective size of population is $N = \frac{1}{40v}, \frac{10}{40v}$ and $\frac{100}{40v}$ in figures 1, 2 and 3, respectively. In each case the solid line represents the least selection $(s = -v/10)$, the broken line selection 10 times as severe (not represented in figure 1 since practically indistinguishable from the preceding) and the dotted line represents selection 100 times as severe.

Figures 4 to 6. Some of the forms taken by the distribution of frequencies of a completely recessive deleterious gene. $Ce^{2Ntq^2}q^{2Nu-1}(1-q)^{4Nu-1}$. Mutation rates are assumed constant and equal $(u = v)$. Effective size of population is $N = \frac{1}{40v}$, $\frac{10}{40v}$ and $\frac{100}{40v}$ in figures 4, 5 and 6, respectively. In each case the solid line represents the least selection $(t = -v/5)$, the broken line selection 10 times as severe (not represented in figure 4 since practically indistinguishable from the preceding) and the dotted line represents selection 100 times as severe.

Figure 7. One of the forms taken by the distribution of gene frequencies in the case in which there is no adaptive difference between the two homozygous types but the heterozygote is selected over both. $Ce^{4Nsq(1-q)}q^{4Nu-1}(1-q)^{4Nu-1}$. In the case represented, $u = v$, $N = \frac{1}{40v}$, $s = 100v$.

Figure 8. The frequencies along two diagonals of the joint distribution for two series of alleles with equal and additive effects on a character on which adverse selection acts according to the square of the deviation from the mean. The solid line shows the frequencies in populations along the line connecting the two favorable types $A_1A_1a_2a_2$ and $a_1a_1A_2A_2$. The broken line refers to the line connecting the extreme types $a_1a_1a_2a_2$ and $A_1A_1A_2A_2$. $\varphi(q) = C[1 - 2s[q_1(1-q_1) + q_2(1-q_2) + 2(q_1 + q_2 - 1)^2]^{2N}q_1^{4Nv_1-1}(1-q_1)^{4Nu_1-1}q_2^{4Nv_2-1}(1-q_2)^{4Nu_2-1}$. In the case shown, $u_1 = v_1 = u_2 = v_2$; $N = 1/2v_1$, $s = 5v_1$. Along the favorable diagonal $q_1 = (1-q_2)$ the distribution is approximately $Ce^{-20q_1(1-q_1)}q_1^2(1-q_1)^2$. Along the unfavorable diagonal $(q_1 = q_2)$ it is approximately $Ce^{-20(1-2q_1(1-q_1))}q_1^2(1-q_1)^2$.

of separate elementary characters, and still less its composition with respect to a single series of alleles. A gene that is favorable in one combination may be deleterious in another. However, if values of w are assigned to each possible combination, the rate of change of the frequency of a particular gene under selection (with specified values of all other gene frequencies) is given by the formula

$$\Delta q_i = \frac{q_i(1 - q_i)}{2\bar{w}} \frac{\partial \bar{w}}{\partial q_i}. \tag{3}$$

If the selection coefficients are small, $\bar{w}$ is close to 1, and the form $\frac{1}{2}q_i(1 - q_i)\partial\bar{w}/\partial q_i$ is sufficiently accurate and is sometimes more convenient.

Two or more of the above factors are usually acting simultaneously. If the rates of change per generation are small, the net rate of change is given sufficiently accurately by the sum.

$$\Delta q = -uq + v(1 - q) - m(q - q_l) + \frac{q(1 - q)}{2} \frac{\partial}{\partial q} \log \bar{w}. \tag{4}$$

Gene frequency is in equilibrium (stable or otherwise) at any point at which $\Delta q = 0$. Opposing mutation pressures, for example, tend to maintain a stable equilibrium at the point $\hat{q} = v/(u + v)$. Mutation opposed by sufficiently strong genic selection [$\Delta q = v(1 - q) - sq(1 - q)$] gives stable equilibrium at the point $\hat{q} = v/s$. Recessive mutation opposed by sufficiently strong selection [$\Delta q = v(1 - q) - sq^2(1 - q)$] gives stable equilibrium at the point $\hat{q} = \sqrt{v/s}$. If there is mutation in both directions and sufficiently strong selection against the heterozygote [$\Delta q = -uq + v(1 - q) - s(1 - 2q)q(1 - q)$] there are two points of stable equilibrium, and one of unstable equilibrium.

Migration pressure, if non-selective, may be written in the same form as mutation pressure, $\Delta q = -m(1 - q_l)q + mq_l(1 - q)$. The theory for migration effects can thus be obtained at any time from that for mutation merely by substituting $m(1 - q_l)$ for u and mq_l for v.

If the population is not indefinitely large, random changes occur in gene frequencies merely as a result of the accidents of sampling among the gametes. Letting N be the effective size of the breeding population, the sample of $2N$ gametes, necessary to replace it, will be distributed according to the expansion of $[(1 - q)A + qA']^{2N}$. The resulting standard deviation of q is $\sqrt{q(1 - q)/2N}$.

The changes in gene frequency due to accidents of sampling are, of course, not correlated in successive generations. Nevertheless, the variance of the probability array for q increases approximately with the number of generations until damped by the approach of q to 0 or 1.

The pressure toward a stable equilibrium point due to mutation, cross breeding and selection, and the divergent tendency due to inbreeding should between them determine a certain *distribution* of values of q which is in equilibrium. The central problem in the genetics of populations is that of finding this distribution under various conditions.

The first attempt at a solution was made by Fisher[3] who used a transformation of scale, $\theta = \cos^{-1}(1 - 2q)$, designed to give a uniform sampling variance, for all values of q. He attempted to express the conditions in a differential equation but reached erroneous conclusions. My first note on the subject was in 1929,[4] the detailed account appearing in 1931.[1] Fisher (1930)[5,6] after inspection of the latter paper in manuscript was able to correct his method so as to yield results in agreement in a number of special cases.

The method followed in the 1931 paper, referred to above, may be summarized as follows. A class of genes with frequency array $[(1 - q)A + qA']$ is distributed in the following generation according to the expansion $[(1 - q - \Delta q)A + (q + \Delta q)A']^{2N}$. The contribution to the class of genes with frequency array $[(1 - q_c)A + q_cA']$ is given by the term in the expansion relating to $2Nq_cA''$s, multiplied by the frequency (f) of the contributing class. The sum of such contributions from all classes of genes should restore the same frequency as in the preceding generation if the form of distribution is in equilibrium. The distinction between gene frequency (q) and frequency (f) of a gene frequency must be kept in mind.

$$f_c = \frac{\lfloor 2N}{\lfloor 2Nq_c \, \lfloor 2N(1 - q_c)} \sum_{q=0}^{1} [(q + \Delta q)^{2Nq_c}(1 - q - \Delta q)^{2N(1 - q_c)}f]. \quad (5)$$

Replacing summation by integration and letting $f = \varphi(q)/2N$ or $\varphi(q)dq$ according to position in the equation, the equation to be solved for $\varphi(q)$ is as follows

$$\varphi(q_c) = \frac{\Gamma(2N)}{q_c(1 - q_c)\Gamma(2Nq_c)\Gamma(2N(1 - q_c))} \int_0^1 (q + \Delta q)^{2Nq_c}(1 - q - \Delta q)^{2N(1 - q_c)}\varphi(q)dq. \quad (6)$$

If Δq is negligibly small—except for enough mutation from the homallelic classes ($q = 0$ or 1) to make equilibrium possible—it may easily be seen that the solution of this equation is

$$\varphi(q) = \frac{C}{q} + \frac{D}{1 - q}. \quad (7)$$

Only the symmetrical case, however, is in complete equilibrium with the homallelic classes, although the rate of change in other cases is extremely slow.

$$\varphi(q) = \frac{(l)}{2(\log 2N + .577)q(1 - q)} \tag{8}$$

when (l) is the proportion of heterallelic loci. The proportions in the homallelic loci are

$$f(0) = \frac{1}{4Nv} f(1/2N),\ f(1) = \frac{1}{4Nu} f\,(1 - 1/2N). \tag{9}$$

$$f(0) + l + f(1) = 1$$

If there is no mutation, there is equilibrium of form among the heterallelic classes when all are approximately equally frequent but falling off at the rate $1/2N$ per generation.

$$\varphi(q) = (l_0 e^{-T/2N}) \tag{10}$$

where l_0 is the initial proportion in heterallelic loci and T is the number of subsequent generations. The two preceding results were confirmed by Fisher[5, 6] by his method.

If mutation is recurring at appreciable rates such that $\Delta q = -uq + v(1 - q)$, the distribution was shown to take the form

$$\varphi(q) = \frac{\Gamma(4Nu + 4Nv)}{\Gamma(4Nu)\Gamma(4Nv)} q^{4Nv-1}(1 - q)^{4Nu-1}. \tag{11}$$

With irreversible mutation, $\Delta q = v(1 - q)$, there is equilibrium of form, with falling off of all class frequencies at the rate v per generation

$$\varphi(q) = (l_0 e^{-vT}) 4Nv q^{4Nv-1}. \tag{12}$$

With genic selection but very small mutation rates, $\Delta q = sq(1 - q)$, equilibrium of form was shown for the distribution

$$\varphi(q) = \frac{Ce^{4Nsq} + D}{q(1 - q)}. \tag{13}$$

The case

$$\frac{Ce^{4Nsq}}{q(1 - q)} \tag{14}$$

is that which is in complete equilibrium with the homallelic classes. Another case is the solution

$$\varphi(q) = \frac{C(e^{4Nsq} - 1)}{q(1 - q)} \tag{15}$$

given by Fisher[5,6] for irreversible mutation.

It was shown finally that the formula for mutation and genic selection combined, $\Delta q = -uq + v(1 - q) + sq(1 - q)$, could be written sufficiently accurately

$$\varphi(q) = Ce^{4Nsq}q^{4Nv-1}(1 - q)^{4Nu-1}. \tag{16}$$

The effects of cross breeding could be introduced in place of (or supplementary to) mutation by the substitution previously referred to.

Haldane[7] has criticized the conclusions derived from these formulae

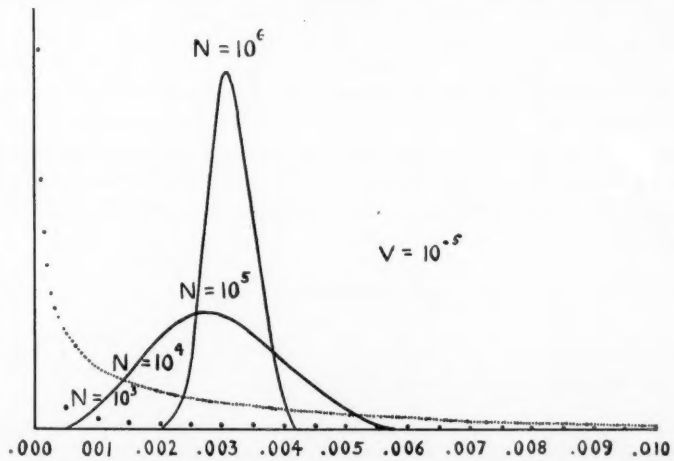

Figure 9. Some of the forms taken by the distribution of frequencies of a recessive lethal gene under different sizes of population. $C(1 - q^2)^{2N}q^{4Nv-1}(1 - q)^{-1}$. The mutation rate ($v$) is taken as 10^{-5}, giving a mean gene frequency of $\sqrt{v} = .0032$ in very large populations. This is approximately realized with $N = 10^6$. With $N = 10^5$, the gene is always present, but $\bar{q}$ is slightly reduced (.0030). In populations with effective size, $N = 10^4$, the gene is absent in about 15%, at any given moment and $\bar{q} = .0020$. With $N = 10^3$, the gene is absent in 87%, and $\bar{q} = .0008$. The case of $N = 10^2$ is not shown as the gene is absent in about 99% and $\bar{q}$ is only .00026. With $N = 10$, the gene is absent in 99.9%, and $\bar{q} = .00008$. In selfed lines ($N = 1$) the gene is absent in 99.996% giving $\bar{q} = .00002$ ($= 2v$).

on the ground that only one type of selection is considered. It is undoubtedly important to extend these results to more general formulations of the action of selection. This has now been done by the same method as above for selection pressure of the form $\Delta q = (s + tq)q(1 - q)$ which applies to any degree of dominance, assuming s and t to be small. I have not been able to reduce the general solution to a simpler form than an infinite series, but in the case of complete equilibrium with the fixed classes there is reduction to the following

$$\varphi(q) = \frac{Ce^{4Nsq + 2Niq^2}}{q(1 - q)}. \qquad (17)$$

Still more general results can be obtained by another approach (presented as an alternative in the preceding paper in the case of mutation pressure). A frequency distribution, ranging from 0 to 1, in which mean and standard deviation do not change under evolutionary pressure and sampling errors, and whose mode (or modes) can be shown to be correct in the limiting case of large population size, must be a good approximation to the desired type. Let $\varphi(q)$ be any distribution of gene frequencies.

Then

$$\bar{q} = \int_0^1 q\varphi(q)dq \text{ (assuming that } \int_0^1 \varphi(q)dq = 1). \qquad (18)$$

Accidents of sampling by themselves can have no effect in changing the mean, the mean of the distribution $[(1 - q)A + qA']^{2N}$ being $2Nq$. The effect of any evolutionary pressure in changing the mean is given by the expression

$$\int_0^1 \Delta q\varphi(q)dq. \qquad (19)$$

If Δq is linear in q, and therefore of the form $\Delta q = -K(q - \hat{q})$ where $\hat{q}$ is the equilibrium point, the change of mean is

$$\int_0^1 \Delta q\varphi(q)dq = -K\int_0^1 (q - \hat{q})\varphi(q)dq = -K(\bar{q} - \hat{q}). \qquad (20)$$

Thus if the mean of $\varphi(q)$ has reached the equilibrium point ($\bar{q} = \hat{q}$) there is no further change. Mutation pressure, $\Delta q = -uq + v(1 - q) = -(u + v)(q - \hat{q})$, comes under this head. The mean of the distribution of gene frequencies is always $\hat{q} = v/(u + v)$ at equilibrium, irrespective of size of population. Cross breeding also comes under this head, $\Delta q = -m(q - q_t)$. Selection pressure on the other hand, with Δq at least a quadratic function of q should give different values of $\bar{q}$ with change in size of population.

The variance of gene frequencies is also easily found in the case of linear evolutionary pressures. The pressure from both directions toward the equilibrium point must be balanced by the scattering effect of accidents of sampling if there is equilibrium in the form of distribution.

$$\sigma_q^2 = \int_0^1 (q - \bar{q})^2\varphi(q)dq = \int_0^1 (q + \Delta q - \bar{q})^2\varphi(q)dq + \int_0^1 \frac{q(1 - q)}{2N}\varphi(q)dq. \qquad (21)$$

Inserting the value $\Delta q = -K(q - \bar{q})$ this reduces to

$$\sigma_q^2 = \frac{\bar{q}(1 - \bar{q})}{1 + 4NK - 2NK^2}. \qquad (22)$$

The term in the denominator involving K^2, tracing to the term in (21) involving $(\Delta q)^2$ is negligible for small values of Δq.

Thus if $\Delta q = -K(q - \bar{q})$, the distribution of gene frequency has the mean, $\bar{q}$; variance, $\sigma_q^2 = \frac{\bar{q}(1 - \bar{q})}{1 + 4NK}$; a range limited at $q = 0$ and $q = 1$ and for large values of N condenses about a single mode at the equilibrium point, $q = \bar{q}$. In Pearson's system of frequency distributions, unimodal curves of limited range come under type I. Assuming a range limited at 0 and 1 and a unit area, the formula for type I is $\varphi(q) = \frac{\Gamma(x + y)}{\Gamma(x)\Gamma(y)} q^{x-1}(1 - q)^{y-1}$. Substituting this value of $\varphi(q)$ in the formulae for $\bar{q}$ and σ_q^2 gives $x = 4NK\bar{q}$, $y = 4NK(1 - \bar{q})$. In the case of mutation, $K = u + v$, $\bar{q} = v/(u + v)$.

$$\varphi(q) = \frac{\Gamma(4Nu + 4Nv)}{\Gamma(4Nu)\Gamma(4Nv)} q^{4Nv-1}(1 - q)^{4Nu-1}. \qquad (23)$$

In the case of migration, $K = m$, $\bar{q} = q_t$

$$\varphi(q) = \frac{\Gamma(4Nm)}{\Gamma(4Nmq_t)\Gamma[4Nm(1 - q_t)]} q^{4Nmq_t-1}(1 - q)^{4Nm(1-q_t)-1}. \qquad (24)$$

These are identical with the formulae derived by the preceding method (cf. 11).

When selection as well as mutation is at work we may write

$$\Delta q = Zq(1 - q) - uq + v(1 - q) \text{ where } Z = {}^1/_2\, d \log \bar{w}/dq. \qquad (25)$$

A suggestion for the distribution formula can be obtained from the special case in which N is so large that there is very little spread from the equilibrium point (or points) at $\Delta q = 0$. The modes should approach the equilibrium points as N is increased. We will make the provisional assumption that the selection coefficients appear in a separate factor from those involving the mutation coefficients and that the desired formula is therefore of the type $\varphi(q) = \psi q^{4Nv-1}(1 - q)^{4Nu-1}$ where ψ is the required function of the selection coefficient. Putting $d \log \varphi(q)/dq = 0$ as a condition for any mode:

$$\frac{d}{dq} \log \varphi(q) = \frac{d}{dq} \log \psi + \frac{4Nv - 1}{q} - \frac{4Nu - 1}{1 - q} = 0. \qquad (26)$$

Thus

$$\frac{q(1-q)}{4N}\frac{d}{dq}\log\psi + v(1-q) - uq + \frac{2q-1}{4N} = 0. \qquad (27)$$

But at the equilibrium point,

$$Zq(1-q) + v(1-q) - uq = 0. \qquad (28)$$

Ignoring the term $(2q-1)/4N$ which tends to disappear as N becomes large

$$Z = \frac{1}{4N}\frac{d}{dq}\log\psi. \qquad (29)$$

$$\psi = Ce^{4N\int Zdq}. \qquad (30)$$

The formula suggested on this basis is thus as follows:

$$\varphi(q) = Ce^{4N\int Zdq}q^{4Nv-1}(1-q)^{4Nu-1}. \qquad (31)$$

This can be written as follows by evaluating $\int Zdq$

$$\varphi(q) = C\bar{w}^{2N}q^{4Nv-1}(1-q)^{4Nu-1}. \qquad (32)$$

We will return to this form later; for the moment it will be convenient to use the following alternative form, easily derived from (25) and (31).

$$\varphi(q) = \frac{Ce^{4N\int\Delta qdq/q(1-q)}}{q(1-q)}. \qquad (33)$$

So far this is merely a suggestion derived from a limiting case. It may be tested, however, for equilibrium in the general case. Testing first for shifting of the mean under evolutionary pressure (see 19)

$$\int_0^1 \Delta q\varphi(q)dq = \frac{C}{4N}\int_0^1 \frac{4N\Delta q}{q(1-q)}e^{\int 4N\Delta qdq/q(1-q)}dq =$$

$$\frac{C}{4N}\Big|_0^1 e^{4N\int Zdq}q^{4Nv}(1-q)^{4Nu}. \qquad (34)$$

Thus there is no shifting of the mean if $4Nv > 0$, $4Nu > 0$.

The test (21) for balancing of the effects on variance of evolutionary pressure and accidents of sampling can be written as follows, omitting the negligible term in $(\Delta q)^2$.

$$2\int_0^1 q\Delta q\varphi(q)dq = -\frac{1}{2N}\int_0^1 q(1-q)\varphi(q)dq. \qquad (35)$$

The left-hand member can be integrated by parts after substituting the suggested value of $\varphi(q)$ and shown to equal the right-hand member.

$$2C\int_0^1 \frac{q\,\Delta q}{q(1-q)}\, e^{4N\int \Delta q dq/q(1-q)}dq =$$

$$\frac{C}{2N}\int_0^1 qd(e^{4N\int \Delta q dq/q(1-q)}) = \frac{C}{2N}\Big|_0^1 e^{4N\int Z dq} q^{4Nv+1}(1-q)^{4Nu} -$$

$$\frac{C}{2N}\int_0^1 e^{4N\int \Delta q dq/q(1-q)}dq = -\frac{1}{2N}\int_0^1 q(1-q)\varphi(q)dq. \qquad (36)$$

The condition of no change in variance is thus met to the same degree of approximation as in the case of mutation. The special case of mutation and genic selection, $\Delta q = -uq + v(1-q) + sq(1-q)$ gives $\varphi(q) = Ce^{4Nsq}q^{4Nv-1}(1-q)^{4Nu-1}$, identical with that (16) derived by the previous more exhaustive method. This holds even in the limiting case (14) in which $\Delta q = sq(1-q)$ but as already noted equilibrium requires that there be some mutation even though the rates are so low that they may be ignored in Δq. The formula (17), obtained by the more exhaustive method in the case of more or less dominance, $\Delta q = (s + tq)q(1-q)$ is also in agreement with the present result.

The effects of certain differences in severity of selection and in effective size of population on the distribution of gene frequencies, assuming no dominance, are illustrated in figures 1 to 3. Figures 4 to 6 make similar comparisons for the case of a completely recessive unfavorable gene. These figures can also illustrate the joint effects of selection, cross breeding and inbreeding in local populations by replacing u and v by $^1/_2m$.

One of the forms taken by the distribution when there is equal selection against both homozygotes in favor of the heterozygote is illustrated in figure 7. With sufficiently smaller population size or sufficiently less intense selection, the distribution would become U-shaped. With a sufficiently larger population size it would become I-shaped about a mean frequency, $q = 0.5$. Increased severity of selection, without increase in size of population, would also pile up the frequencies about this point.

In a large population, mutation opposed by moderately severe selection tends to hold the deleterious gene at a low frequency, $\hat{q} = v/s$ if semi-dominant, $\hat{q} = (v/s)^{1/2}$ if completely recessive. In a sufficiently small sample from such a population, selection of the same degree of severity becomes ineffective and the mean frequency of the deleterious gene gradually rises to the equilibrium point due to opposing mutation pressures $\hat{q} = v/(u+v)$. If unfavorable mutation is much more frequent than the reverse ($v > u$) this may mean a shift to approximate fixation of the deleterious gene. The rate of approach to the new mean, after a reduction in size of population is, however, extremely slow, being dependent on mutation pressure.

In the case of a deleterious *recessive* factor, the immediate effect of reduction in size of population is indeed the reverse of that indicated above, though the final effect, a rise in mean frequency of the deleterious gene toward the point $q = v/(u + v)$ occurs at length in this case as well as when dominance is lacking (provided that populations in which the deleterious gene is fixed can persist). The immediate decrease in frequency on reduction of size of population is illustrated in figure 9 in the extreme case of a recessive lethal in which case there can, of course, be no secondary rise in frequency. Taking the dominant as the type, $\bar{w} = 1 - q^2$, $\Delta q = v(1 - q) - q^2/(1 + q)$ giving the approximate equilibrium point $q = v^{1/2}$. The distribution is

$$\varphi(q) = C(1 - q^2)^{2N} q^{4Nv - 1}(1 - q)^{-1}. \tag{37}$$

The mean, $\bar{q}$, can easily be expressed in Γ functions (on substituting x for q^2). It reduces approximately to $\bar{q} = \dfrac{\Gamma(2Nv + 1/2)}{\sqrt{2N}\,\Gamma(2Nv)}$. For values of $2Nv$ larger than 1, this is close to $v^{1/2}$ i.e., gene frequency varies about the equilibrium point. If on the other hand $2Nv$ is a small fraction, q is approximately $v(2\pi N)^{1/2}$ which means a great reduction in frequency of lethals in populations as the effective size of inbreeding units falls below $1/2v$.

So far we have derived formulae for distributions only where the selection coefficients are constant. But as already noted, it is really the system of gene frequencies that is more or less adaptive, not the isolated genes. No adequate picture of the evolutionary process can be made without taking factor interaction into account.

The momentary selection pressure in cases of factor interaction was given in (3), giving as the distribution of gene frequencies

$$\varphi(q_i) = C\bar{w}^{2N} q_i^{4Nv_i - 1}(1 - q_i)^{4Nu_i - 1}. \tag{38}$$

Here $\bar{w}$ is the mean selective value with q_i variable, but a specified set of values for the other gene frequencies.

The joint frequency surface must be such that on assigning specified values to all of the q's but one, the distribution for that one is that given by (38). The joint distribution

$$\varphi(q_1, q_2, \ldots q_n) = C\bar{w}^{2N} \Pi_{i=1}^{n} q_i^{4Nv_i - 1}(1 - q_i)^{4Nu_i - 1} \tag{39}$$

when $\bar{w}$ is the mean selective value in terms of all of the q's as variables, satisfies this condition and is thus the desired form.

As an example consider the case of a character for which the grade of development depends on the additive effects of multiple factors, lacking dominance, but for which the selective value falls off as the square of its deviation from an optimum. The selective value has been shown to be as follows:[2]

$$\bar{w} = 1 - K[2\Sigma\alpha_i^2 q_i(1 - q_i) + (M - O)^2]$$

where α_i is the effect of gene A_i with frequency q_i; $M(= 2\Sigma\alpha_i q_i)$ is the mean and O is the optimal grade.

The multidimensional surface, $\bar{w}$, has in general many peaks, separated by shallow saddles. The shallowest saddles are those between optimal combinations differing only in two pairs of factors, e.g., $A_1A_1a_2a_2\ldots\ldots$ and $a_1a_1A_2A_2\ldots\ldots$

The nature of the distribution along and across such a saddle is illustrated in figure 8 in the case of two pairs of factors with equal effects. With smaller N the distribution from $A_1A_1a_2a_2$ to $a_1a_1A_2A_2$ becomes U-shaped. With larger N or weaker selection it becomes I-shaped about $q = 0.5$. Stronger selection pushes the modes toward the favored homallelic types.

The evolutionary implications will not be discussed here in detail. For the most part the present results merely put the conclusions previously reached[1,8,9] on a more definite basis. These conclusions may be summarized briefly as follows.

In large freely interbreeding populations with no secular change in conditions of life for long periods of time, all gene frequencies approach equilibrium at a certain peak $\bar{w}$, not necessarily the highest peak. Under secular change in conditions the surface $\bar{w}$ itself changes and there is evolutionary change in the system of gene frequencies, following the changes in position of the controlling peak. Evolution here may be said to be guided by intragroup selection.

In sufficiently small completely isolated populations, the random divergencies of gene frequencies from their equilibrium values become important, tending to bring about approximate fixation of some random combination of genes which is not likely to be a peak combination. The result is a largely nonadaptive differentiation. In extreme cases there may be the deterioration which characteristically follows excessive inbreeding. Isolation may here be considered the dominating evolutionary factor.

In a large population subdivided into numerous small, partially isolated groups, the combination of directed and random divergencies in gene frequencies, associated with intergroup selection, gives a trial and error mechanism under which the system of gene frequencies may pass from lower to higher peak values of $\bar{w}$ and the species may evolve continuously even without secular changes in conditions (although this process, occurring in all species, itself tends to bring about such secular changes). The combination of partial isolation of subgroups with intergroup selection seems to provide the most favorable conditions for evolutionary advance.

Mutation is always a factor in providing material for evolution but may be said to dominate the course of evolution only in so far as mutants

appear which are fertile *inter se* but largely infertile with the parent type, i.e., when mutation is itself an isolating factor.

[1] Wright, S., *Genetics*, **16**, 97–159 (1931).
[2] Wright, S., *Jour. Genetics*, **30**, 243–256 (1934).
[3] Fisher, R. A., *Proc. Roy. Soc. Edinburgh*, **42**, 321–341 (1922).
[4] Wright, S., *Amer. Nat.*, **63**, 556–561 (1929).
[5] Fisher, R. A., *The Genetical Theory of Natural Selection*, Oxford, Clarendon Press, 272 pp. (1930).
[6] Fisher, R. A., *Proc. Roy. Soc. Edinburgh*, **50**, 205–220 (1930).
[7] Haldane, J. B. S., *The Causes of Evolution*, New York and London, Harper Brothers, 235 pp. (1932).
[8] Wright, S., *Proc. Sixth Internat. Congress of Genetics*, **1**, 356–366 (1932).
[9] Wright, S., *Jour. Genetics*, **30**, 257–266 (1935).

THE INFLUENCE OF WAVE-LENGTH ON GENETIC EFFECTS OF X-RAYS[1]

BY HUGO FRICKE AND M. DEMEREC

THE WALTER B. JAMES LABORATORY FOR BIOPHYSICS OF THE LONG ISLAND BIOLOGICAL ASSOCIATION AND THE DEPARTMENT OF GENETICS OF THE CARNEGIE INSTITUTION OF WASHINGTON, COLD SPRING HARBOR, N. Y.

Communicated May 13, 1937

A great deal of work has been done with genetic effects of x-ray radiation. It has been shown that this effect, as measured by lethals produced in mature sperm of *Drosophila melanogaster*, is proportional to the dosage in r-units. It has also been found that the dosage-genetic effect relationship is not affected by the wave-length within a range of 0.02 to 2.0 Å.

It is a problem of theoretical importance to determine whether or not the mechanism which produces genetic effects is dependent upon wave-length, especially within the range of soft rays. In studying this range, however, technical difficulties due to high absorption are encountered.

The purpose of this work was to obtain data on the genetic effect of soft rays in experiments where both physical and biological sides were well controlled. A particular effort was made to control the absorption. While these experiments were in progress, results of similar experiments conducted by Timofeeff-Ressovsky and Zimmer[2] were published. Results of both experiments agree, although there is a disagreement in their interpretation.

Experimental Procedure. Physical Part.—The x-rays were obtained from a tungsten tube with a bulb of lithium glass having a 0.02 mm. window for the exit of the soft rays. The conditions of irradiation are given in table 1. The quality of the rays was determined by measuring their ab-

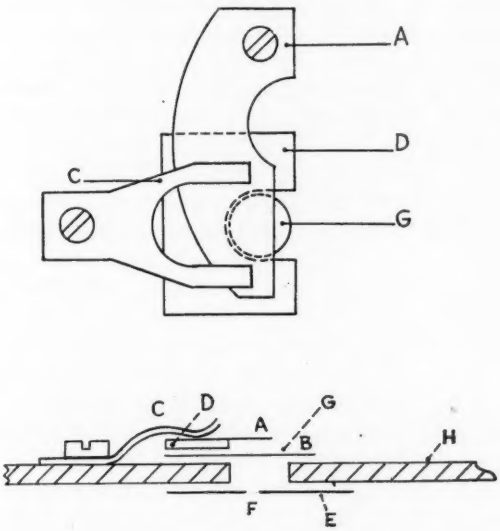

FIGURE 1

Diagrammatic representation of the device used for compressing abdomen of the male during irradiation. Insect is placed at *G* between celluloid sheets *A* and *B* and compressed to about 0.15 millimeters.

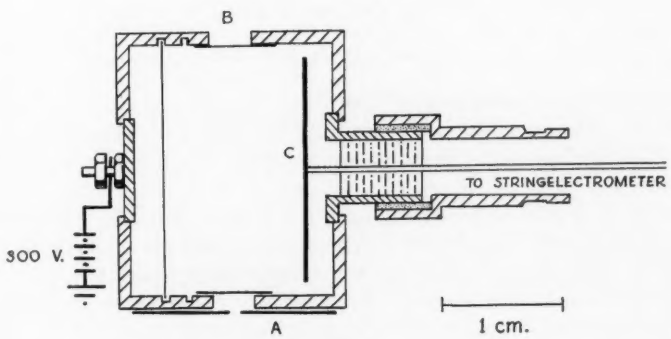

FIGURE 2

Diagram of the small ionization chamber used for measuring x-ray intensity.

sorption in aluminium, the effective wave-length (λ eff.) being the wave-length of homogeneous radiation which is absorbed in the aluminium to the same extent as the rays tested. The thickness of the aluminium is such as to decrease the x-ray intensity 50 per cent. The degree of homogeneity of the rays is shown by the absorption curves for celluloid in figure 3.

During the irradiation the abdomen of the insect was compressed to about 0.15 mm. by the device shown schematically in figure 1. The insect was anesthetized with ether and, with the aid of a low-power microscope, placed between the celluloid foils *A* and *B*, each of which was 0.127 mm. in

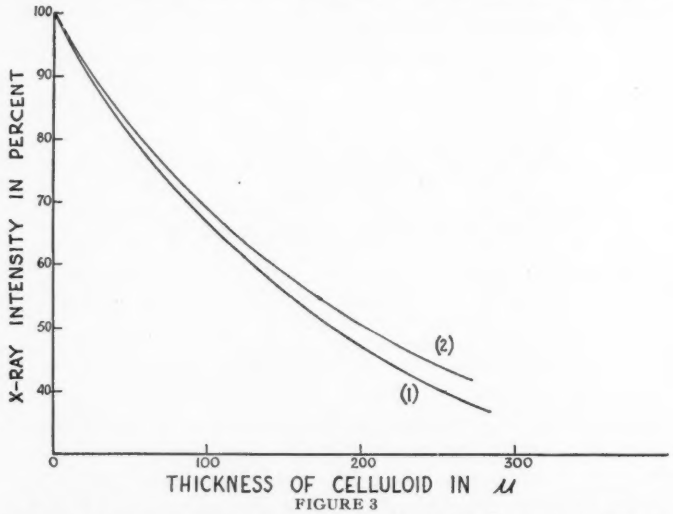

FIGURE 3

Curves representing penetration of x-rays (λ eff. 2.2 Å) through different thicknesses of celluloid. (1) and (2) refer to rays which have passed through filters of 0.127 and 0.254 millimeters of celluloid, respectively.

thickness, pressure being exerted with the phosphorbronze spring *C*. The degree of compression was determined by the thickness of the aluminium foil *D*. *B* is supported by mounting over a hole (a few mm. in diameter) in the brass plate *H*.

The determination of the x-ray intensity in r/sec.was made by means of an ionization chamber similar to the one previously described.[3] This chamber was larger than required for the present range of wave-length and since the high absorption of the rays in the air space of the chamber made it somewhat inconvenient to use, the actual measurements were made with a smaller chamber constructed as shown in figure 2. This chamber consisted of a cylindrical metal box with the apertures *A* and *B* covered with cellophane 0.02 mm. thick, for the entrance and exit of the rays. The electrode *C*,

connected to a string electrometer, collected the ions from the volume given by the interior depth of the chamber multiplied by the area of the entrance aperture. Tests in which these two dimensions were changed showed the ionization (when corrected for the absorption of the rays in the cellophane and in the air of the chamber) to be proportional to this volume, for the range of wave-lengths used. The comparison with the standard chamber showed that the chambers agreed to within 2 per cent.

The distance of the insect from the target of the x-ray tube was 15 cm. The aperture in the x-ray housing was adjusted to prevent the rays from striking any of the metal parts of the insect holder, in order to avoid the emission of secondary radiations which could reach the insect. After one-half of the x-ray dosage had been applied the insect holder was reversed and the remaining treatment given through the opposite side, thus insuring a nearly homogeneous irradiation of the insect.

In order to measure the incident x-ray intensity, the insect holder was removed and the ionization chamber placed with its front aperture in place of the insect; the 0.127 mm. celluloid was placed in the path of the rays during this measurement. In the intensity thus measured a slight correction factor had to be introduced for the absorption of the rays in the cellophane which covers the front aperture of the chamber and in the air space extending from the front aperture to the center of the chamber.

Next we had to determine the extent by which the x-ray intensity was decreased in passing through the insect. The ionization chamber was mounted immediately behind the insect holder, and the diaphragm E with the aperture F (Fig. 1) placed between them; the size of the aperture was chosen so that it was just sufficient to allow all rays passing through the abdomen of the insect to enter the ionization chamber. The diaphragm was mounted on the insect holder and its position adjusted under the microscope in such a way that the aperture became centered in relation to the insect. The x-ray intensities I_0 and I_1 obtained without and with the insect in place, were recorded. The ratio $\alpha' = I_1/I_0$ gives an approximate value for the permeability of the insect to the rays, but is slightly greater than the true value α because the hardness of the rays increased by passing through the celluloid B (the hardening of the rays due to their absorption in the ionization chamber can be neglected). The average values of α' were found to be $\lambda = 0.94$ Å, $\alpha' = 0.936$; $\lambda = 2.2$ Å, $\alpha' = 0.631$. For $\lambda = 0.94$ Å the difference between α and α' is negligible and the average intensity of the rays in the insect was obtained by multiplying the incident intensity by $\sqrt{\alpha'}$.

For $\lambda = 2.2$ Å we proceeded as follows. Figure 3 shows the penetration of the rays through various thicknesses of celluloid; curve (1) was obtained with celluloid A interposed as a filter in the path of the rays, thus giving a picture of the penetration of the rays through the insect, and curve (2) was

obtained with celluloid B used as an additional filter. Curve 2 shows that the permeability, $\alpha' = 0.631$, measured above, corresponds to 0.131 mm. of celluloid. The insect therefore had, on the average, the same permeability as this thickness of celluloid. We consequently obtain the average intensity of the rays in the insect by averaging the ordinates of curve 1 between 0 and 0.131 mm. celluloid and multiplying this value (0.777) by the incident intensity.

Biological Part.—Males from inbred wild type Florida stock were used in these experiments. Treated sperm was tested for induced X-chromosome lethals by *ClB* method. All lethals found were tested through another generation to confirm that they were lethal and also to determine, through linkage relationship, whether or not any of them were connected with a chromosomal aberration. The breeding procedure was as follows: (1) *ClB/ec ct v g* ♀ × 1♂ + Florida treated; (2) F_1 − *ClB*/(+ treated) × *ec ct v g;* (3) F_2 − flies from cultures without males were tested further to confirm that lack of males was due to an induced lethal and to determine, through linkage relationship between *ec*, *ct*, *v* and *g*, if any chromosomal aberration involving the X-chromosome was induced. Treated males were kept with females for six days following the treatment and were then discarded. This was done to insure that only sperm which was mature at the time of the treatment was used in tests.

Experimental Results.—Individual treatment was given to each male, the dosage for each treatment was determined by measurements of incident and transmitted intensities, as explained above, and the biological effect induced by the treatment was analyzed. The dosages applied at $\lambda = 2.2$ Å ranged from 1330 to 1420 r units, the average being 1376 r units; those applied at $\lambda = 0.94$ Å ranged from 1302 to 1395 r units with an average of 1354 r units. The results of the biological tests are summarized in table 1.

TABLE 1

PHYSICAL AND BIOLOGICAL DATA ON TREATMENTS OF *Drosophila melanogaster* MALES WITH X-RAYS OF DIFFERENT WAVE-LENGTH

(C. A. = chromosomal aberration)

POTENTIAL KV.	CURRENT M. A.	FILTER MM. AL	λ EFFECTIVE Å	AVERAGE DOSAGE	NUMBER OF ♂'S TREATED	NUMBER OF ♂'S FERTILE	CLB/ + CULTURES TOTAL	STERILE	FERTILE	LETHALS TOTAL	LETHALS PER CENT	LETHALS TESTED FOR C. A.	C. A.	C. A. PER CENT
18	25	0.50	0.94	1354	101	71	1930	481	1449	52	3.59	42	7	19
8	25	0	2.2	1376	101	66	1928	498	1430	50	3.50	40	5	12
Control	..	....	....	....	...	..	....	...	2108	23	1.14	23	0	..

The data show that a high proportion of treated males did not produce any offspring, which is not surprising in view of the rough handling these males received. About 25 per cent of F_1 females were sterile. An almost identi-

cal percentage of X-chromosome lethals was observed for the two wave-lengths used. This frequency was about three times as high as was found in the controls which were raised at the same time. As already stated, inbred Florida wild type stock was used for both the treatments and for the control. This stock possesses a factor which increases the natural rate of mutations and which is responsible for the high frequency of lethals found in the controls.[4]

Although the data on the frequency of chromosomal aberrations are not large, they indicate that there is no difference between the material treated with these two wave-lengths.

Discussion.—It may be assumed that the genetic effect considered here originates as chemical changes in genes, which are induced (either directly or through intermediate reactions) by the activations which in turn are produced by the absorption of the rays. For constant dosage, the number of activations may be considered to be independent of the wave-length,[5] so any dependence of genetic effects upon wave-length would have to be explained with reference to the special distribution of the activations along the path of the photoelectrons.[6] The statistical chance that any particular atom in the cell is activated is about 4×10^{-8} for a dosage of 1000 r. This value is much smaller than the values obtained for the chance of inducing specific genetic changes. For example, the rate of change from wild to white locus is about 1 in 10,000 for 1000 r of dosage.[7] There are three ways in which this situation may be explained.

1. We may assume that a gene contains approximately 2500 atoms, and that the activation of any of these atoms induces the chemical change required to produce the genetic effect. This would indicate an average gene diameter of about 25 Å. As the wave-length of x-rays increases, the distance between adjacent activated atoms decreases. When the distance becomes of the same order of magnitude as the gene diameter, we should expect to find that gene changes are dependent on wave-length. For $\lambda = 2.2$ Å the distance between adjacent activations is about 25 Å, and therefore the fact that at such high values of the wave-length the number of gene changes due to x-rays was independent of the wave-length would indicate that gene changes are not produced by direct activations inside the gene.

2. The change in the gene is in the nature of a sensitized reaction induced by the transfer of energy from neighboring activated molecules. Studies of chemical reactions obtained by the x-irradiation of solutions of organic chemicals[8] have shown the importance of sensitized transformations of organic molecules induced by the activation of the water molecule. It has been possible to demonstrate[9] that the activated water molecules may travel over a considerable distance (of the order of 10^{-6} cm.), encountering numerous collisions with the organic molecules, before reaction takes place.

It may, therefore, be assumed (with respect to the occurrences in the cell) that the primary chemical transformations, resulting from the passage of a photoelectron, take place in a comparatively large space around the path of the photoelectron. This makes the original relative positions of the activated molecules unimportant and thus makes the resulting effects independent of the wave-length. With this assumption, the chance that the gene is acted upon involves its competition with other cell constituents for the energy of the activated molecules. It may be worth while to derive some kind of measure of this interference of the cell constituents. Assume that each x-ray activation causes a chemical transformation of one atomic group, that the nucleus contains N atomic groups which are chemically acted upon and that the probability that any one of them is changed, is the same for all of them. The number of activations may be taken as 2×10^{18} per cc. and per 1000 r. The volume of the nucleus is taken as 10^{-11} cc. The number of activations inside the nucleus is therefore 20,000. Using 10^{-4} as the chance that any particular one of the N atomic groups is acted upon (for a dosage of 1000 r) it follows that $N = 2 \times 10^8$. To see the significance of this value, assume that the nucleus contains 20 per cent organic material. If the average molecular weight of the organic constituents was 6000, then the number of organic molecules in the nucleus would be $200/(6000) \times 6 \times 10^{23} \times 10^{-14} = 2 \times 10^8$, which is the value found for N. Expressed in another way: Taking the average atomic weight of the non-aqueous phase of the nucleus as 12, the number of reactive groups interfering with the action of the x-rays on the gene, is one for each 500 atoms of the organic constituents.

3. In the chemical changes produced by the x-irradiation of the cell, we may differentiate between two phases: (*a*) the extremely rapid primary reactions which are induced by the initial activations as discussed in 2, and (*b*) slower secondary reactions (involving the whole medium of the cell) induced by the primary changes. One might inquire whether the gene changes may not be produced during this second phase [viz., be induced by the chemical change in the surrounding medium rather than during the first phase as assumed under 2]. The evidence obtained from studies of the influence of wave-length does not allow any differentiation between these two possibilities.

Summary.—Drosophila melanogaster males were treated with x-rays of wave-lengths 0.94 Å and 2.2 Å and the frequency of induced X-chromosome lethals was determined by genetic methods. Males were treated individually in a specially designed chamber where their abdomens were compressed to a thickness of about 0.15 millimeters for the purpose of reducing the absorption of these soft rays. The method used in measuring the dosage (in r units) is described.

The average frequency of lethals among the sperm obtained from males

treated with an average dosage of 1354 r units at 0.94 Å is 3.59 per cent; while the frequency of lethals among the sperm of males treated with 1376 r units at 2.2 Å is 3.50 per cent. The frequency of lethals per unit dose is, within the experimental accuracy, the same for both wave-lengths.

It is pointed out that this evidence indicates that these genetic changes are not produced by direct activations within a gene but they are probably in the nature of sensitized reactions induced by the transfer of energy from neighboring activated molecules.

[1] This work has been supported by a grant to Hugo Fricke by the Radiation Committee of the National Research Council. We are indebted to Dr. M. T. Jones for his assistance in the physical measurements.

[2] Timofeeff-Ressovsky, N. W., and Zimmer, K. G., *Strahlentherapie*, **54**, 265–278 (1935).

[3] Fricke, H., and Morse, S., *Phil. Mag.*, **7**, 129 (1929).

[4] Demerec, M., *Genetics*, **22**, 190 (1937).

[5] The dosage in r can, with sufficient approximation, be considered a measure of the x-ray energy absorbed by the cell.

[6] Fricke, H., and Petersen, B. W., *Am. Jour. Roentg. and Radium Ther.*, **17**, 611–620 (1927).

[7] Demerec, M., *Cold Spring Harbor Symposia Quant. Biol.*, **3**, 110–115 (1934).

[8] Fricke, H., and Petersen, B. W., *Am. Jour. Roentg. and Radium Ther.*, **17**, 611–620 (1927).

Fricke, H., and Morse, S., *Ibid.*, **18**, 426–430 (1927).

Fricke, H., and Brownscombe, E. R., *Jour. Am. Chem. Soc.*, **55**, 2358–2363 (1933); *Phys. Rev.*, **44**, 240 (1933).

Fricke, H., *Jour. Chem. Phys.*, **2**, 556–557 (1934).

Fricke, H., and Hart, E. J., *Ibid.*, **3**, 60, 364–365 (1935); *Ibid.*, **4**, 418–422 (1936).

Stenstrom, W., and Lohmann, A., *Jour. Biol. Chem.*, **79**, 673–678 (1928); *Jour. Rad. Soc. Am.*, **16**, 322–327 (1931).

Stenstrom, W., *Jour. Rad. Soc. Am.*, **13**, 437–440 (1929); *Ibid.*, **17**, 432 (1931).

Clark, G. L., and Pickett, L. W., *Jour. Am. Chem. Soc.*, **52**, 465–479 (1930).

Clark, G. L., Pickett, L. W., and Johnson, E. D., *Jour. Rad. Soc. Am.*, **15**, 245–261 (1930).

Clark, G. L., and Fitch, K. R., *Ibid.*, **17**, 285–293 (1931).

Clark, G. L., and Coe, W. S., *Jour. Chem. Phys.*, **5**, 97–105 (1937).

[9] Fricke, H., Hart, E. J., and Smith, H. P.—unpublished work.

ON THE THEORY OF BEAMS RESTING ON A YIELDING FOUNDATION

BY M. ERIC REISSNER

DEPARTMENT OF MATHEMATICS, MASSACHUSETTS INSTITUTE OF TECHNOLOGY

Communicated May 12, 1937

Introduction.—The purpose of this note is to advance the theory of the following problem: An elastic beam of infinite length, loaded inside a finite region, rests on a yielding foundation. It is asked how much the results of this theory are influenced by alterations of the assumptions concerning the physical nature of the foundation. As it is known, in engineering the assumption is made that at every point the deflection is proportional to the foundation pressure at this point and independent of the pressure at points distant from the considered point. This assumption, which is justified for a beam resting on water, has the advantage of leading to the solution of a linear differential equation. In reality the deflection at one point depends almost always on the pressure distribution along the whole beam and the question leads, as it will be seen, to an integro-differential equation.

On this problem a very interesting paper has recently been published by M. A. Biot.[1] There the foundation is represented by a semi-infinite elastic body. Here it shall be shown that the formal results of that paper can be obtained in a different, more general way; more general in that it is shown to be possible to solve the problem explicitly for every kind of foundation whether elastic or having properties more difficult to describe, provided the deflection of the surface of the foundation due to a concentrated loading is known.

It will also be shown that under this assumption the problem of buckling of a beam of infinite length, resting on a yielding foundation and on an infinity of equidistant supports, can be solved completely. So far as known a treatment of this problem has been given only for the beam resting on water.[2]

Statement of the Problem.—The differential equation for the deflection w of an elastic beam of constant stiffness E_bJ is given by

$$E_bJ\frac{d^4w}{dx^4} = p(x). \tag{1}$$

There $p(x)$ denotes the acting load. Supposing that the beam rests on a yielding foundation, p consists of the given load p_0 and the foundation pressure p_u

$$p = p_0 - p_u. \tag{2}$$

The deflection of the surface of the foundation at a point x, due to a loading $p_u(\xi)\, d\xi$ inside a region $(\xi, \xi + d\xi)$ may be denoted by $dv(x)$. Then

$$dv(x) = K(|x - \xi|)\, p_u(\xi)\, d\xi \tag{3}$$

where K is a function depending on the nature of the foundation which may be determined experimentally or otherwise.

The deflection $v(x)$ due to the whole foundation pressure is then

$$v(x) = \int_{-\infty}^{\infty} K(|x - \xi|)\, p_u(\xi)\, d\xi. \tag{4}$$

For reasons of continuity $v(x)$ is equal to the beam deflection $w(x)$. Furthermore p_u is connected with w and the loading p_0 by equations (1) and (2). Thus from (4) the following integro-differential equation for w is obtained.

$$w(x) = \int_{-\infty}^{\infty} K(|x - \xi|)\left[p_0(\xi) - E_b J \frac{d^4 w}{d\xi^4}\right] d\xi \tag{5}$$

or differentiating both sides with respect to x

$$\frac{dw}{dx} = \int_{-\infty}^{\infty} \frac{\partial K(|x - \xi|)}{\partial x}\left[p_0(\xi) - E_b J \frac{d^4 w}{d\xi^4}\right] d\xi. \tag{6}$$

Equations (5) and (6) are the fundamental equations of the problem of the elastically supported beam of infinite length. It is possible to give an explicit solution of (6) without knowing K, provided K has certain properties which are always fulfilled for physical reasons and which do not need to be stated here.

It may be remarked that working with equation (6) instead of (5) is suggested by the fact that for the case of a foundation consisting of a *two*-dimensional semi-infinite elastic body K becomes infinite for large values of its argument and hence also w, whereas its derivatives stay finite.

The Solution.—Assuming for the sake of simplicity an even load distribution $p_0(x) = p_0(|x|)$ we write

$$p_0(x) = \int_0^{\infty} P(\lambda) \cos \lambda x d\lambda \tag{7}$$

where $P(\lambda)$ is given by

$$P(\lambda) = \frac{1}{\pi} \int^{\infty} p_0(x) \cos \lambda x dx. \tag{8}$$

Then we try a solution in the form

$$\left.\begin{aligned} w(x) &= \int_0^\infty W(\lambda)\cos\lambda x d\lambda \\ \frac{dw}{dx} &= -\int_0^\infty \lambda W(\lambda)\sin\lambda x d\lambda \\ \frac{d^4w}{dx^4} &= \int_0^\infty \lambda^4 W(\lambda)\cos\lambda x d\lambda \end{aligned}\right\} \tag{9}$$

Introducing (7) and (9) into (6) gives

$$-\int_0^\infty \lambda W(\lambda)\sin\lambda x d\lambda = \int_0^\infty \frac{\partial K(|x-\xi|)}{\partial x} \left\{\int_0^\infty P(\lambda)\cos\lambda\xi d\lambda - E_b J\int_0^\infty \lambda^4 W(\lambda)\cos\lambda\xi d\lambda\right\} d\xi \tag{10}$$

and changing the order of integration in the right side of (10), a process which is easily justified,

$$\int_0^\infty \lambda W(\lambda)\sin\lambda x d\lambda = \int_0^\infty [E_b J\lambda^4 W(\lambda) - P(\lambda)] \int_{-\infty}^\infty \frac{\partial K(|x-\xi|)}{\partial x}\cos\lambda\xi d\xi d\lambda. \tag{11}$$

Putting $\xi - x = u$ (12)

$$\int_{-\infty}^\infty \frac{\partial K(|x-\xi|)}{\partial x}\cos\lambda\xi d\xi = \cos\lambda x\int_{-\infty}^\infty \frac{\partial K(|u|)}{\partial u}\cos\lambda u du - \sin\lambda x\int_{-\infty}^\infty \frac{\partial K(|u|)}{\partial u}\sin\lambda u du. \tag{13}$$

Because $\frac{\partial K(|u|)}{\partial u}$ is an odd function the first term on the right in (13) vanishes and we have

$$\int_{-\infty}^\infty \frac{\partial K(|x-\xi|)}{\partial x}\cos\lambda\xi d\xi = -\sin\lambda x\int_{-\infty}^\infty \frac{\partial K(|u|)}{\partial u}\sin\lambda u du. \tag{14}$$

Placing

$$\int_{-\infty}^\infty \frac{\partial K(|u|)}{\partial u}\sin\lambda u du = k(\lambda) \tag{15}$$

(11) can be written

$$\int_0^\infty \lambda W(\lambda) \left[1 + E_b J \lambda^3 k(\lambda) \right] \sin \lambda x d\lambda = \int_0^\infty P(\lambda) k(\lambda) \sin \lambda x d\lambda \qquad (16)$$

This implies

$$\lambda W(\lambda) = \frac{P(\lambda) k(\lambda)}{1 + E_b J \lambda^3 k(\lambda)} \qquad (17)$$

and hence

$$\frac{dw}{dx} = - \int_0^\infty \frac{P(\lambda) k(\lambda)}{1 + E_b J \lambda^3 k(\lambda)} \sin \lambda x d\lambda. \qquad (18)$$

Thus provided K and consequently the Fourier transform k of its derivative is known, the solution of (6) is explicitly given by (18).

As an application we may take the case of a two-dimensional semi-infinite elastic foundation of the same breadth $2b$ as the beam, where the result is already known.[1] Here[3]

$$K(|x - \xi|) = \frac{1}{\pi b E} \ln |x - \xi| + \text{const.} \qquad (19)$$

and consequently

$$\frac{\partial K(|u|)}{\partial u} = \frac{1}{\pi b E u}. \qquad (20)$$

Hence from (15)

$$k(\lambda) = \frac{1}{\pi b E} \int_{-\infty}^{\infty} \frac{\sin \lambda u}{u} du = \frac{1}{bE} \qquad (21)$$

and from (18)

$$\frac{dw}{dx} = - \frac{1}{bE} \int_0^\infty \frac{P(\lambda)}{1 + \frac{E_b J}{bE} \lambda^3} \sin \lambda x d\lambda. \qquad (22)$$

The Stability Problem.—When the beam is compressed in its plane, the loading p_0 is given by the stress component vertical to the plane which arises from a deflection. If the stress resultant is equal to $P_0 kg$ this component is

$$p_0 = - P_0 \frac{d^2 w}{dx^2} \qquad (23)$$

and (6) becomes

$$\frac{dw}{dx} = -\int_{-\infty}^{\infty} \frac{\partial K(|\, x - \xi \,|)}{\partial x} \left[P_0 \frac{d^2w}{d\xi^2} + E_b J \frac{d^4w}{d\xi^4} \right] d\xi. \qquad (24)$$

A solution of this equation can be obtained which satisfies the following conditions

$$w(x_n) = 0, \left(\frac{d^2w}{dx^2}\right)_{x = x_n} = 0; \quad x_n = \pm nl, \; n = 0, \; 1, \; 2, \ldots. \qquad (25)$$

These conditions state that the beam is freely supported at the points x_n. Trying

$$w = \sin \frac{m\pi x}{l}, \quad m = 1, \; 2, \; 3, \ldots. \qquad (26)$$

(25) is satisfied and (24) becomes

$$\frac{m\pi}{l} \cos \frac{m\pi x}{l} = \int_{-\infty}^{\infty} \frac{\partial K(|\, x - \xi \,|)}{\partial x} \left[P_0 \left(\frac{m\pi}{l}\right)^2 - E_b J \left(\frac{m\pi}{l}\right)^4 \right] \sin \frac{m\pi\xi}{l} \, d\xi \qquad (27)$$

$$\cos \frac{m\pi x}{l} = \frac{m\pi}{l} \left[P_0 - E_b J \left(\frac{m\pi}{l}\right)^2 \right] \cos \frac{m\pi x}{l} \int_{-\infty}^{\infty} \frac{\partial K(|\, u \,|)}{\partial u} \sin \frac{m\pi u}{l} \, du \qquad (28)$$

and with (15)

$$1 = \frac{m\pi}{l} \left[P_0 - E_b J \left(\frac{m\pi}{l}\right)^2 \right] k \left(\frac{m\pi}{l}\right) . \qquad (29)$$

Hence the critical stress resultant is given by

$$P_0 = \frac{1}{\frac{m\pi}{l} k \left(\frac{m\pi}{l}\right)} + E_b J \left(\frac{m\pi}{l}\right)^2. \qquad (30)$$

With the help of this result the question which number m of half waves yields the smallest critical P_0 can be answered. It is obtained from

$$\frac{\partial P_0}{\partial m} = \frac{l}{\pi} \frac{d}{dm} \left(\frac{1}{mk \left(\frac{m\pi}{l}\right)} \right) + 2m \left(\frac{\pi}{l}\right)^2 E_b J = 0. \qquad (31)$$

For the example of the two-dimensional elastic foundation[2] (21), (30) is

$$P_0 = E_b J \left(\frac{m\pi}{l}\right)^2 + \frac{bEl}{\pi m} \tag{32}$$

and from (31)

$$m_{\min} = \frac{1}{\pi} \sqrt[3]{\frac{Ebl^3}{2E_b J}}, \tag{33}$$

whereas in the case of a beam resting on water it is[2]

$$P_0 = \left(\frac{m\pi}{l}\right)^2 E_b J + \frac{\eta l^2}{m^2\pi^2}; \; m_{\min} = \frac{1}{\pi} \sqrt[4]{\frac{\eta l^4}{E_b J}}. \tag{34}$$

Concluding Remarks.—From the foregoing lines it can be seen that the problem of a beam resting on a yielding foundation admits a simple solution if the beam is of infinite length.

The same problem is by far more complicated for a beam of finite length. Here some results have been obtained by Wieghardt.[4] The present author would like to add that he has recently made some progress in this direction which he hopes to publish soon.

[1] Biot, M. A., *Jour. Appl. Mech.*, **4**, 1–8 (1937).
[2] Timoshenko, St., *Theory of Elastic Stability*, McGraw-Hill, 108–112 (1936).
[3] Timoshenko, St., *Theory of Elasticity*, McGraw-Hill, 88 (1934).
[4] Wieghardt, K., *Zeit. ang. Math. Mech.*, **2**, 165–184 (1922).

GROUPS HAVING A MAXIMUM NUMBER OF INDEPENDENT GENERATORS

BY G. A. MILLER

DEPARTMENT OF MATHEMATICS, UNIVERSITY OF ILLINOIS

Communicated April 30, 1937

It is obvious that the number of operators in a set of independent generators of a given group G cannot exceed the number of the prime factors of the order of G. When G has a set of independent generators involving as many operators as there are prime factors in its order we shall say that it has a maximum number set of independent generators. All the operators of such a set are of prime order, for if one of them would be of composite order the cyclic group generated by it could be successively extended by the other operators in the given set of independent generators. The group

obtained by each such extension is such that its order involves at least one prime factor which does not appear in the order of the preceding group and this prime factor divides the order of G. The extending operator could therefore have been selected so that it appears in a Sylow subgroup of G. Hence there results the following theorem: *If a group has a maximum number set of independent generators then the operators of the set can be so selected that each of them appears in a Sylow subgroup of the group and is of prime order.*

When the operators of a maximum number set of independent generators of G have been so selected that each of them appears in a Sylow subgroup of G it results that the operators of the same order in such a set are relatively commutative and that every Sylow subgroup of G is abelian and of type 1^k. To prove this statement it may first be noted that such a set contains at least one set of independent generators of the Sylow subgroup of G whose order is the smallest prime number which divides the order of G and that these generate an abelian subgroup of type 1^k contained in G since two such independent generators of G generate a group whose order is at least equal to the product of their orders. When the order of G exceeds the order of this Sylow subgroup every other operator of the given set of independent generators together with this Sylow subgroup generates a subgroup of G whose order is the order of this Sylow subgroup multiplied by a prime number.

If this independent generator was selected in such a way that this prime number is as small as possible it may be replaced by an operator whose order is equal to this prime number in case that its order is less than this number. If the order of G is divisible by the square of this number the same process may be repeated until we arrive at a subgroup of G which is generated by two Sylow subgroups of G. The Sylow subgroup whose order is a power of the larger of these two prime numbers is invariant under this subgroup of G. It is abelian and its type is of the form 1^k in view of the number of independent generators of G. If the order of G is divisible by more than two prime numbers this process can be repeated and hence there results the following theorem: *A necessary condition that a group has a set of independent generators which involves as many operators as there are prime factors in its order is that all its Sylow subgroups are abelian and of type* 1^k *and that one of these Sylow subgroups is invariant and gives rise to a quotient group which has a maximum number set of independent generators.*

It should be noted that some groups which have sets of independent generators which involve as many operators as there are prime factors in their orders, have also sets of independent generators which separately involve a smaller number of operators. This is the case, for instance, of every cyclic group whose order is not divisible by the square of a prime number but is composite. In a group whose order is a power of a prime

number it is well known that the number of operators in a set of independent generators is the same for every possible such set.[1] An elementary example of a case where this is also true with regard to a group whose order is not a power of a prime number and where every possible set of independent generators is a maximum number set is furnished by extending the abelian group of order p^m and of type 1^m, p being an odd prime number, by an operator of prime order less than p which transforms each of its operators into the same power of itself.

In a given maximum number set of independent generators of G each of the operators of highest order generates a subgroup which is invariant under all the operators of G but not all of the operators of the Sylow subgroup whose order is the largest prime number which divides the order of G need to generate separately invariant subgroups under G. This is illustrated by the group of order 36 obtained by extending the non-cyclic group of order 9 by two operators of order 2 which transform respectively all the operators of this non-cyclic group into their inverses and only two of these operators of order 3 into their inverses. Similarly, each of the operators in the given set of independent generators which generate a Sylow subgroup whose order is a power of another prime number generates a subgroup which is invariant under all the operators of the given set of independent generators whose orders are a smaller prime number. *These conditions are clearly sufficient as well as necessary in order that* G *may contain a maximum number set of independent generators.*

It has been noted that the only prime power group which has a maximum number set of independent generators is the abelian group whose type is of the form 1^k. From this it results directly that there is one and only one abelian group of an arbitrary order which has the property that it has a maximum number set of independent generators, viz., the direct product of abelian groups of prime power order which are separately of type 1^k. While there is one and only one abelian group of a given order which has a maximum number set of independent generators the number of non-abelian groups of a given order which have this property depends on the order and may vary from zero to an arbitrarily large number. This is illustrated by the fact that m such non-abelian groups of order $2p^m$, p being an odd prime number, can be constructed as follows: Start with the abelian group of order p^m and of type 1^m, and extend this group successively by operators of order 2 such that the first transforms each of the independent generators of this abelian group into its inverse, the second transforms one of these m generators into itself and each of the others into its inverse, the third transforms two of them into themselves and each of the others into its inverse, etc. Each of these groups except the first has also a set of independent generators which involves less than $m + 1$ operators.

It results from the given properties of G that every group which has a maximum number set of independent generators is solvable. It was noted above that the Sylow subgroup whose order is a power of the largest prime number which divided the order of G is invariant under G. If the Sylow subgroup whose order is a power of the smallest prime number which divides the order of G is also invariant under G then G is the direct product of this Sylow subgroup and of the group generated by its other Sylow subgroups. This results from the well-known theorem that if two invariant subgroups have only the identity in common then the group contains the direct product of these two invariant subgroups. When the order of G is the product of two and only two prime numbers there is obviously only one group which does not have a maximum number set of independent generators, viz., the cyclic group whose order is the square of a prime number, but only one of the five groups whose order is the cube of a prime number has a maximum number set of independent generators.

If the order of G is of the form pq^n, p and q being prime numbers such that $p - 1$ is divisible by q, two and only two of the possible groups for an arbitrary value of n satisfy the condition that each of them contains a maximum number set of independent generators. One of these is abelian and the other is non-abelian. This results directly from the fact that the group of isomorphisms of the invariant subgroup of order p is cyclic and hence involves only one subgroup of order q. When $p - 1$ is not divisible by q there is only one such group. Whenever $n > 1$ a set of independent generators of such a group can be so selected that it is composed also of n operators, and G contains invariantly the direct product of its subgroup of order p and the abelian groups of order q^{n-1} and of type 1^{n-1}. When it is non-abelian all of its operators except the $p - 1$ of order p are of order q. In particular, it results from these theorems that there are two—and only two—groups of each of the orders 12, 24, 48, 96, etc., which separately have a set of independent generators involving as many operators as there are prime factors in their orders.

While every group which contains a set of independent generators involving as many operators as there are prime factors in its order is solvable a group which involves a set of independent generators composed of one less operator is not necessarily solvable. This results from the fact that the icosahedral group involves a set of three independent generators. In fact, any set of three operators of order 2, which generate the icosahedral group, is such a set of independent generators since every two of these operators generate a dihedral subgroup. An illustration of a group which satisfies a necessary condition but not a sufficient condition, as noted above, that a group has a maximum number set of independent generators, is furnished by the non-abelian group of order 75 obtained by extending the non-cyclic group of order 25 by an operator of order 3 which transforms it into itself.

This group contains 50 operators of order 3 because the extending operator of order 3 could not transform into itself any of its operators of order 5.

[1] Miller, Blichfeldt, Dickson, *Finite Groups*, 1916, p. 127.

TRIHORNOMETRY: A NEW CHAPTER OF CONFORMAL GEOMETRY

BY EDWARD KASNER

DEPARTMENT OF MATHEMATICS, COLUMBIA UNIVERSITY

Read before the Academy, April 26, 1937

I. *Introduction and Definitions.*—The object of this paper is to study the relationships between the conformal invariants of an ordered triplet of curves C_1, C_2 and C_3, which pass through a common point in a common direction. This configuration is termed a *trihorn*. We shall consider only the case where the curves have different curvatures at the common point so that there are no contacts of higher order. A trihorn has six fundamental invariants and these are connected by inequalities and equalities, of which the most important are the formulae (5), (6), (8), (9), (11) and (13). In particular we obtain the analogues of the laws of sines and cosines of ordinary trigonometry.

Let x represent the curvature and y the derivative of curvature with respect to length of arc of a curve C of the trihorn at the common point. The ordered pair of curves (C_i, C_j) of our trihorn form a horn angle and the unique absolute conformal invariant

$$M_{ij} = \frac{(x_j - x_i)^2}{y_j - y_i} \tag{1}$$

is called the *measure of the horn angle.* It is noted that the measure is never zero, but it may be infinite. When the measure is infinite, then the horn angle is said to be circular.

The quantity

$$\alpha_{ij} = \frac{\dfrac{y_i - y_k}{x_i - x_k}}{\dfrac{y_j - y_k}{x_j - x_k}} = \frac{(y_i - y_k)(x_j - x_k)}{(x_i - x_k)(y_j - y_k)} \tag{2}$$

is also an absolute conformal invariant. It is to be regarded as a new kind of angle, namely, the angle between the horn (C_i, C_k) and the horn (C_j, C_k). This is called a *dihorn angle.*

We note the obvious relations

$$M_{ji} = -M_{ij}, \alpha_{ji} = \frac{1}{\alpha_{ij}}.$$

All the invariants are of third order since they involve curvatures and rates of variation of curvatures. (There is also, for three curves, an absolute invariant of second order

$$\Gamma_{1,23} = \frac{x_3 - x_1}{x_2 - x_1}$$

which I shall not use in this paper. This is in fact invariant under all point transformations.)

We thus have six fundamental parts in a trihorn, the three horns (or sides or M's) and the three dihorns (or angles or α's). The following theorems are stated without proof.

II. *Discussion of Special Trihorns.*—THEOREM 1. *If two measures of a trihorn are infinite, then all the measures are infinite and the angles are indeterminate. This trihorn is said to be circular.*

THEOREM 2. *If one angle of a trihorn is indeterminate, then all the angles are indeterminate and all the measures are infinite. Thus the trihorn is circular.*

THEOREM 3. *If only* $M_{ij} = \infty$, *then* $M_{jk}M_{ki} < 0$, $\alpha_{jk} = 0$, $\alpha_{ki} = \infty$ *and* $\alpha_{ij} \neq 0, 1, \infty$. *This is called a partially circular trihorn.*

THEOREM 4. *If* $\alpha_{jk} = 0$ (*or* $\alpha_{ki} = \infty$), *then* $\alpha_{ki} = \infty$ (*or* $\alpha_{jk} = 0$), $\alpha_{ij} \neq 0, 1, \infty$, $M_{ij} = \infty$ *and* $M_{jk}M_{ki} < 0$. *Thus the trihorn is partially circular.*

THEOREM 5. *For the partially circular trihorn of theorems* 3 *and* 4, *we have the unique equality*

$$\alpha^2_{ij} M_{ki} + M_{jk} = 0. \tag{3}$$

A trihorn which is neither circular nor partially circular is termed a non-circular trihorn. For a non-circular trihorn, the measures and the angles are determinate, finite, non-zero numbers.

If a non-circular trihorn is such that

$$M_{12} + M_{23} + M_{31} = 0, \tag{4}$$

then it is said to be *wide-open*. If a non-circular trihorn is not wide-open, then it is termed a general trihorn.

THEOREM 6. *If a non-circular trihorn is wide-open, then each angle is unity. Conversely if an angle of a non-circular trihorn is unity, then all the angles are unity and the trihorn is wide-open.*

If a non-circular trihorn is such that $M_{ij} + M_{jk} = 0$, then it is termed an isosceles trihorn. Obviously no wide-open trihorn can be isosceles, and conversely.

III. *Solutions of General Trihorns.*—If we are given any three parts (except the three angles) of a general trihorn, we proceed to show, in the remaining six theorems, how to find the other three parts. Several of the cases turn out to be *ambiguous*, namely, there may be two solutions or one solution.

In the following, by a *number* we shall mean a determinate, finite, non-zero number. Moreover, when we say that a number is an angle of a trihorn, we shall understand that the number also is distinct from unity.

THEOREM 7. *For three numbers* $\mathbf{M_{12}}$, $\mathbf{M_{23}}$, $\mathbf{M_{31}}$, *to be the measures of a general non-isosceles trihorn, it is necessary and sufficient that no sum* $\mathbf{M_{ij} + M_{jk}}$ *be zero, and that*

$$M_{12}M_{23}M_{31}\,(M_{12} + M_{23} + M_{31}) < 0. \tag{5}$$

Then all the angles have two distinct values which are given by the formulae

$$\alpha_{ij} = \frac{-M_{jk}M_{ki} \pm \sqrt{-M_{ij}M_{jk}M_{ki}\,(M_{ij} + M_{jk} + M_{ki})}}{M_{ki}\,(M_{ij} + M_{ki})}. \tag{6}$$

THEOREM 8. *For three numbers* $\mathbf{M_{12}}$, $\mathbf{M_{23}}$, $\mathbf{M_{31}}$, *to be the measures of a general isosceles trihorn, it is necessary and sufficient that a sum* $\mathbf{M_{ij} + M_{jk}}$ *be zero, and that all the measures be not equal in absolute value. Then the angles are uniquely determined by the formulae*

$$\alpha_{ij} = \frac{2M_{ij}}{M_{ij} + M_{ki}},\quad \alpha_{jk} = \frac{M_{ij} - M_{ki}}{2M_{ij}},\quad \alpha_{ki} = \frac{M_{ij} + M_{ki}}{M_{ij} - M_{ki}}. \tag{7}$$

THEOREM 9. *For three numbers* α_{12}, α_{23}, α_{31}, *to be the angles of a general trihorn, it is necessary and sufficient that*

$$\alpha_{12}\alpha_{23}\alpha_{31} = 1. \tag{8}$$

Then the ratios of the measures are uniquely determined by the formulae

$$\frac{M_{12}}{\alpha_{31}(1 - \alpha_{12})} = \frac{M_{23}}{\alpha_{12}(1 - \alpha_{23})} = \frac{M_{31}}{\alpha_{23}(1 - \alpha_{31})}. \tag{9}$$

This is the analogue of the law of sines in ordinary trigonometry.

THEOREM 10. *The necessary and sufficient conditions that the numbers* $\mathbf{M_{ij}}$, $\mathbf{M_{ki}}$ *and* α_{jk} *belong to a general trihorn are*

$$M_{ij}\alpha_{jk} + M_{ki} \neq 0,\ M_{ij}\alpha_{ik}^2 + M_{ki} \neq 0. \tag{10}$$

The remaining parts are uniquely determined by the formulae

$$M_{jk} = -\frac{(M_{ij}\alpha_{jk} + M_{ki})^2}{M_{ij}\alpha_{jk}^2 + M_{ki}}$$

$$\alpha_{ij} = \frac{M_{ij}\alpha_{jk} + M_{ki}}{M_{ij}\alpha_{jk}^2 + M_{ki}} \tag{11}$$

$$\alpha_{ki} = \frac{1}{\alpha_{ij}\alpha_{jk}}.$$

The first formula of (11) *is the analogue of the law of cosines of ordinary trigonometry.*

In the above theorem we are given two sides and the included angle; in the next theorem we are given two sides and an adjacent angle.

THEOREM 11. *For three numbers* M_{ij}, M_{jk} *and* α_{jk} *to be the measures and angle of a general trihorn, it is necessary and sufficient that*

(*a*) If $M_{ij} + M_{jk} \neq 0$, then

$$M_{jk}^2 + 4M_{ij}M_{jk}\alpha_{jk}(1 - \alpha_{jk}) \geqq 0. \tag{12}$$

The remaining parts have two possible determinations which are given by the formulae

$$\left.\begin{aligned} M_{ki} &= \frac{-2M_{ij}\alpha_{jk} - M_{jk} \pm \sqrt{M_{jk}^2 + 4M_{ij}M_{jk}\alpha_{jk}(1 - \alpha_{jk})}}{2}, \\ \alpha_{ij} &= \frac{M_{ij}\alpha_{jk} + M_{ki}}{M_{ij}\alpha_{jk}^2 + M_{ki}}, \qquad \alpha_{ki} = \frac{1}{\alpha_{ij}\alpha_{jk}}. \end{aligned}\right\} \tag{13}$$

(*b*) If $M_{ij} + M_{jk} = 0$, then $\alpha_{jk} \neq -1, {}^1/_2$. The remaining parts are uniquely determined by the formulae

$$\left.\begin{aligned} M_{ki} &= -2M_{ij}\alpha_{jk} - M_{jk} \\ \alpha_{ij} &= \frac{M_{ij}\alpha_{jk} + M_{ki}}{M_{ij}\alpha_{jk}^2 + M_{ki}}, \quad \alpha_{ki} = \frac{1}{\alpha_{ij}\alpha_{jk}}. \end{aligned}\right\} \tag{14}$$

THEOREM 12. *If any three numbers are given, a trihorn can be constructed with two of the numbers as angles and the third as measure. The remaining parts are found by formulae* (8) *and* (9).

We observe that the *special metric* considered in this paper is based on the distance element $ds = dx^2/dy$, that is, on the calculus of variation problem

$$\int \frac{dx^2}{dy} = \int \frac{1}{y'}\,dx = \text{extremum}.$$

It is therefore non-riemannien, and of course non-euclidean.

The fundamental "congruence group" is

$$X = mx + a, \quad Y = m^2y + b;$$

it is induced by the conformal group of the plane $w = f(z)$, and shows the effect on differential elements of third order. The extremals are straight lines (wide-open phenomenon), and transversality is defined by taking half the slope. The angle α of a "right angle" is therefore $\alpha = \frac{1}{2}$. The reciprocal $\alpha = 2$ defines anti-perpendicularity.

With relation to the above integral, angle (dihorn) is properly defined as

$$A = \frac{1}{2} \log \alpha;$$

so that instead of (8) we have, in any trihorn,

$$A_{12} + A_{23} + A_{31} = 0, A_{ji} = -A_{ij};$$

but we have preferred to use α so as to make all our formulas algebraic. For related material on conformal geometry see our earlier papers: *Proc. Intern. Congr. Math.*, **2,** 81 (1912); *Proc. Nat. Acad. Sci.* (with G. Comenetz), **22,** 303 (1936); *Science*, **85,** 480 (1937).

I wish to thank J. de Cicco and S. Gorn for assistance in writing the present paper.

For horn angles of all orders of contact we obtain a non-archimedean geometry. For second order contact the fundamental fifth order invariant includes Mullin's inversive invariant as a special case.

CONTINUOUS RINGS AND THEIR ARITHMETICS

BY J. v. NEUMANN

INSTITUTE FOR ADVANCED STUDY, PRINCETON, NEW JERSEY

Communicated May 20, 1937

Introduction. 1. This note continues the analysis of the geometrical systems called *continuous geometries*, discussed in four previous notes of the author.[1] Again only results and outlines of proofs will be given, the details being reserved for subsequent publications.[2]

The main result obtained in A. T. was this: Every continuous geometry L is (lattice) isomorphic to the principal right-ideal lattice $R_{\mathfrak{R}}$ of a suitable regular ring $\mathfrak{R}$, which is uniquely determined by L, up to a (ring) isomorphism.[3] This result establishes a complete characterization (up to a lattice isomorphism) of L, by means of the purely algebraic entity $\mathfrak{R}$, and is satisfactorily complete under this aspect. Yet it is incomplete in so far as it fails to characterize those regular rings $\mathfrak{R}$, which arise in this manner

from continuous geometries L. Indeed, the $R_{\mathfrak{R}}$ of a regular ring need not be a continuous geometry—it is in general merely a *complemented, modular lattice* (R. R., 710), and conversely, the above quoted result holds for all complemented, modular lattices L, subject only to a mild restriction.[4]

Hence the following problem arises:

(*) Characterize those regular rings $\mathfrak{R}$, for which the principal right-ideal lattice $R_{\mathfrak{R}}$ is a continuous geometry. They will be called *continuous rings*. Since the characterization of a continuous geometry is self dual (cf. C. G., 94, *Remark* 2, 94–96, 99), and since $R_{\mathfrak{R}}$ and $L_{\mathfrak{R}}$ are dual (anti-isomorphic, cf. R. R., 710), so both $R_{\mathfrak{R}}$ and $L_{\mathfrak{R}}$ are continuous geometries, if one is. Hence (*) is really right-left symmetric.

It follows from the above quoted result, that

(**) L isomorphic to $R_{\mathfrak{R}}$ establishes a (up to isomorphisms) one-to-one correspondence between all continuous geometries L and all continuous rings $\mathfrak{R}$.

In this note we will give a complete algebraic characterization of the continuous rings. We shall also see that those rings are the infinite limiting case of the *simple rings*. (§§2–4) An analysis of the notions of algebraicity and transcendency (with respect to the center), which permits applications to the theory of elementary divisors, follows (§§5–7). Finally the first steps toward an arithmetic of continuous rings are made (§8). These latter investigations reveal quite unexpected conditions: They permit to extend the notion of an integer from algebraic to transcendental elements (in a continuous ring), and they show that continuous rings are in many ways simpler than discrete ones (i.e., matrix rings over division algebras; cf. the end of §7).

The Rank. 2. Consider first a system L satisfying our axioms I–VI (C. G., 94–96), which therefore is either a discrete projective geometry of $n-1$ dimensions: $L = L_n$, $n = 1,2, \ldots$, or a continuous geometry $L = L_\infty$ (E. C. G., 101). Assume that it is isomorphic to the $R_{\mathfrak{R}}$ of a regular ring $\mathfrak{R}$—which is certainly the case for $L = L_n$, $n = 4,5, \ldots, \infty$ (cf. A. T., 20, and[4]). We identify L with the (isomorphic) $R_{\mathfrak{R}}$, and the dual geometry L' with $L_{\mathfrak{R}}$ (which is dual to $R_{\mathfrak{R}}$, cf. R. R., 710). We will denote the dimension functions of L and L' by $D(\mathfrak{a})$, $\mathfrak{a} \epsilon L$, and $D'(\mathfrak{a}')$, $\mathfrak{a}' \epsilon L'$, respectively.

We can prove:

(1) For every $a \epsilon \mathfrak{R}$

$$D((a)_r) = D'((a)_l) = 1 - D((a)_l^r) = 1 - D'((a)_r^l).^5$$

We therefore define:

(2) Denote the common value of the four quantities enumerated in (1) by $R(a)$, and call it the *rank of* a.

The range of $R(a)$ is the same as the range of $D(\mathfrak{a})$, that is, the set $D_n =$

$\left(0, \frac{1}{n}, \frac{2}{n} \ldots, 1\right)$ for $L = L_n$, $n = 1,2, \ldots$,[6] and the set D_∞ of all real numbers $\geqq 0$, $\leqq 1$ for $L = L_\infty$ (C. G. 98, E. C. G. 101).

A simple analysis discloses that the rank $R(a)$ possesses the following properties:[7]

(3) (α) Always $0 \leqq R(a) \leqq 1$.
(β) $R(a) = 0$ if and only if $a = 0$.
(γ) $R(a) = 1$ if and only if a^{-1} exists.
(δ) $R(a) = R(b)$ if and only if $a = ubv$, where u^{-1}, v^{-1} exist.
(ϵ) $R(ab) \leqq \text{Min}\,(R(a), R(b))$.
(η) $R(a + b) \leqq R(a) + R(b)$.
(θ) For $e^2 = e, f^2 = f, ef = fe = 0$ we have $R(e + f) = R(e) + R(f)$.

Also:

(4) Even the following requirements, which are weakened forms of (3), (α)–(θ), possess the only solution $\overline{R}(a) \equiv R(a)$:

($\bar{\alpha}$) Always $0 \leqq \overline{R}(a) \leqq 1$.
($\bar{\beta}$) $\overline{R}(0) = 0$.
($\bar{\gamma}$) $\overline{R}(1) = 1$.
($\bar{\epsilon}_r$) $\overline{R}(ab) \leqq \overline{R}(a)$.
($\bar{\epsilon}_l$) $\overline{R}(ab) \leqq \overline{R}(b)$.
($\bar{\theta}$) For $e^2 = e, f^2 = f, ef = fe = 0$ we have $\overline{R}(e + f) = \overline{R}(e) + \overline{R}(f)$.

It suffices to require only one of the two conditions ($\bar{\epsilon}_r$) and ($\bar{\epsilon}_l$). Since (4) refers only to the ring operations in $\mathfrak{R}$, and since it is right-left symmetric, we can infer:

(5) Every (ring) automorphism or anti-automorphism of $\mathfrak{R}$ leaves $R(a)$ invariant.

3. We define the *rank-distance* of two $a, b \in \mathfrak{R}$ to be $R(a - b)$. One derives immediately from (3) that this is a *metric* in $\mathfrak{R}$:

(i) $R(a - b) \begin{cases} = 0 \text{ for } a = b \\ > 0 \text{ for } a \neq b, \end{cases}$
(ii) $R(a - b) = R(b - a)$,
(iii) $R(a - c) \leqq R(a - b) + R(b - c)$,

and that $a + b$, $a \cdot b$ fulfil Lipschitz conditions:

(iv) $R((a + b) - (c + d)) \leqq R(a - c) + R(b - d)$,
(v) $R(ab - cd) \leqq R(a - c) + R(b - d)$.

For $L = L_n$, $n = 1,2, \ldots$, $R(a - b)$ does not lead to an interesting topology in $\mathfrak{R}$ because its values are $0, \frac{1}{n}, \frac{2}{n} \ldots, 1$ only—but for $L = L_\infty$, where

its values are all real numbers $\geqq 0$, $\leqq 1$, it becomes topologically significant.

$\mathfrak{R}$ is *complete* in the topology of the rank distance, that is, we can prove this:

(6) If $a_1, a_2, \ldots \epsilon \mathfrak{R}$, then the existence of an $a \epsilon \mathfrak{R}$ with

$$\lim_{\rho \to \infty} R(a_\rho - a) = 0,$$

which will be denoted by

$$\lim_{\rho \to \infty}{}^{(R)} (a_\rho) = a,$$

is equivalent to

$$\lim_{\rho, \sigma \to \infty} R(a_\rho - a_\sigma) = 0.$$

The topology of the rank distance gives us also a means to characterize the principal right (left) ideals among all right (left) ideals of $\mathfrak{R}$ (cf. R. R., 707 [Def. 1]). We can show:

(7) A right (left) ideal of $\mathfrak{R}$ is a principal one (and hence an element of $R_{\mathfrak{R}} = L$ $(L_{\mathfrak{R}} = L')$) if and only if it is a closed set in the topology of the rank distance.

The really significant case for (6), (7) is of course the continuous one: $L = L_\infty$. In the discrete cases $L = L_n$, $n = 1,2, \ldots$, this topology degenerates, and so (6) becomes vacuous, while (7) states that then all right (left) ideals are principal ones (cf. A. T., 18).

Characterization of the Continuous Rings. 4. Let us start at the opposite end. Let a regular ring $\mathfrak{R}$ be given. We assume that it is irreducible, that is, that its center $\mathfrak{Z}$ is a division algebra. (Cf. R. R., 712, the $\mathfrak{R}$ of our §§2–3 was irreducible along with L.) We define:

(8) $\mathfrak{R}$ is a *rank ring*, if a numerical function $\overline{R}(a)$ can be defined for all $a \epsilon \mathfrak{R}$, which possesses the properties (4), $(\bar{\alpha})$–$(\bar{\theta})$, with the following provisos: $(\bar{\beta})$ has to be replaced by the stronger condition

$$(\bar{\bar{\beta}}) \quad \overline{R}(a) = 0 \text{ if and only if } a = 0.$$

Both conditions $(\bar{\epsilon}_r)$ and $(\bar{\epsilon}_l)$ must be required. We can show that the requirements of (8) imply also the equivalent of (3) (η):

$$(\bar{\eta}) \quad \overline{R}(a + b) \leqq \overline{R}(a) + \overline{R}(b).$$

This permits us to infer that $\overline{R}(a - b)$, which we will again call a *rank distance*, is a *metric* in $\mathfrak{R}$: That is, that (i)–(iii) (at the beginning of §3) hold. Therefore we can define further:

(9) The $\mathfrak{R}$ of (8) is *complete*, if it is complete in the topology of a rank distance $\overline{R}(a - b)$, as described in (6).

We are now able to formulate our main result:

(10) For a regular (and irreducible) ring $\mathfrak{R}$ the principal right ideal lattice $R_{\mathfrak{R}}$ (and also its dual, the principal left-ideal lattice $L_{\mathfrak{R}}$) fulfils our axioms I–VI (C. G., 94–96) if and only if $\mathfrak{R}$ is a *complete rank ring*.

(11) If this is the case, then the rank $\overline{R}(a)$ of (8) is unique, and coincides with the rank $R(a)$ of (2). Hence all statements of §§2–3 hold for it. Now we see that $L = R_{\mathfrak{R}}$ is either an $L = L_n$, $n = 1,2, \ldots$, or $L = L_\infty$. In the first case $\mathfrak{R}$ is an nth order matrix algebra over a suitable division algebra $\mathfrak{d}$ (cf.[6]): $\mathfrak{R} = \mathfrak{d}_n$, and we call $\mathfrak{R}$ a *discrete ring*. In the second case $\mathfrak{R}$ is a *continuous ring*, as defined in (*) in §1.

The discrete rings $\mathfrak{R}$ coincide, by one of Wedderburn's famous theorems,[8] with all *simple rings*. Hence our notion of a continuous ring is the infinite limiting case of Wedderburn's notion of a simple ring.

Algebraic and Transcendental Elements. 5. From now on $\mathfrak{R}$ will denote a fixed complete rank ring. Let again be $L = R_{\mathfrak{R}}$, $L' = L_{\mathfrak{R}}$, $R(a)$ the (unique) rank. Later on we will even require that this $\mathfrak{R}$ be a continuous ring.

For a discrete $\mathfrak{R}$, that is $\mathfrak{R} = \mathfrak{d}_n$ at most n^2 elements of $\mathfrak{R}$ can be linearly independent with respect to $\mathfrak{d}$; hence for any $a \epsilon \mathfrak{R}$ a linear relation, with coefficients from $\mathfrak{d}$, must exist between the powers $1, a, a^2, \ldots$ of a:

(π) $p(a) = 0$, where $p(x) \equiv x^l + \alpha_1 x^{l-1} + \ldots + \alpha_l$,

(π_1) $\alpha_1, \ldots, \alpha_l \epsilon \mathfrak{d}$.

If $\mathfrak{d}$ has a finite (linear) order over its own center, which is also the center $\mathfrak{Z}$ of $\mathfrak{R}$, then the same argument applies to $\mathfrak{Z}$ instead of $\mathfrak{d}$, that is, every $a \epsilon \mathfrak{R}$ fulfils even an equation (π) with

(π_2) $\alpha_1, \ldots, \alpha_l \epsilon \mathfrak{Z}$.

But for an arbitrary $\mathfrak{d}$ this need not be true.

Observe that (π), (π_2) leads to a much more satisfactory theory than (π), (π_1): The familiar theory of polynomial equations applies in the first case only, since only $\mathfrak{Z}$ is commutative.

In a continuous $\mathfrak{R}$ this entire argument breaks down, since it contains any number of linearly independent elements. It is also discouraging that only $\mathfrak{Z}$ (the center of $\mathfrak{R}$), but no $\mathfrak{d}$, can be defined for a continuous $\mathfrak{R}$ (cf. A. T., 21–22, §7)—yet only (π), (π_1) is generally true, and it involves $\mathfrak{d}$.

6. In spite of these unfavorable symptoms, we will base our discussion in a continuous $\mathfrak{R}$ on its center $\mathfrak{Z}$. Let P be the set of all polynomials

(P) $p(x) \equiv x^l + \alpha_1 x^{l-1} + \ldots + \alpha_l$, $\alpha_1, \ldots, \alpha_l \epsilon \mathfrak{Z}$.

If $a \epsilon \mathfrak{R}$, then a $p(x) \epsilon P$ with $p(a) = 0$ need not exist. Let us therefore investigate how near we can get to $p(a) = 0$, that is, how small we can make $R(p(a))$. The most unfavorable case covers those $p(x) \epsilon P$, for which

(a) $R(p(a)) = 1$, that is $p(a)^{-1}$ exists. The other $p(x) \epsilon P$ have

(b) $R(p(a)) < 1$, that is $p(a)^{-1}$ does not exist, those we call *a-singular*.

Among them the irreducible ones (with respect to the coefficient domain $\mathfrak{Z}$) are particularly important. Define:

(12) $\mathfrak{Z}$ is the set of all irreducible, a-singular polynomials from P.

Several relations between a, P, $\mathfrak{Z}$ hold both for discrete and continuous $\mathfrak{R}$ and can be established with little difficulty. We enumerate them:

(13) $\mathfrak{Z}$ is enumerable.[9] Denote its elements by $q_1(x)$, $q_2(x)$, ...

(14) Let α be the *gr.l.b.* of all $R(p(a))$, for all $p(x)\epsilon P$.[10] Let $\mathfrak{a}(\mathfrak{a}')$ be the intersection of all $(p(a))_r$, $(((p(a))_l)$, for all $p(x)\epsilon P$. Then there exists a unique *ip.* $e\epsilon\mathfrak{R}$ with

$$(e)_r = \mathfrak{a},\ (e)_l = \mathfrak{a}',$$

We have $R(e) = D(\mathfrak{a}) = D'(\mathfrak{a}') = \alpha$.

(15) Define α_i, $\mathfrak{a}_i$, $\mathfrak{a}'_i$ $(i = 1,2, \ldots)$ as the α, $\mathfrak{a}$, $\mathfrak{a}'$ in (14), but restricting $p(x)$ to the $(q_i(x))^t$, $t = 1,2, \ldots$. Then there exists a unique *ip.* $e_i\epsilon\mathfrak{R}$ with

$$(e_i)_r = \mathfrak{a}_i,\ (e_i)_l = \mathfrak{a}'_i).$$

We have $R(e_i) = D(\mathfrak{a}_i) = D'(\mathfrak{a}'_i) = \alpha_i$.

Put $\beta_i = 1 - \alpha_i$, $f_i = 1 - e_i$; so $R(f_i) = \beta_i$.

(16) We have $ef_i = f_ie = 0$, $f_if_j = 0$ for $i \neq j$, and $e + \Sigma f_i = 1$.[11] Also $\alpha + \sum_i \beta_i = 1$. Hence the rings $\mathfrak{R}(e)$, $\mathfrak{R}(f_i)$ $(i = 1,2, \ldots)$ (cf. R. R., 713) are mutually orthogonal.

(17) e and all f_i commute with every x which commutes with a, hence in particular with a. For every such x a unique decomposition

$$x = x_e + x_{1-e} = x_e + \sum_i x_{fi},^{11}\ x_e\epsilon\mathfrak{R}(e),\ x_{1-e}\epsilon\mathfrak{R}(1 - e),\ x_{fi}\epsilon\mathfrak{R}(f_i)$$

exists.

(18) For every $p(x)\epsilon P$ the $p(a_e)$, when formed in $\mathfrak{R}(e)$, possesses a $p(a_e)^{-1}$ in $\mathfrak{R}(e)$.

(19) Form $p(a_{1-e})$ in $\mathfrak{R}(1 - e)$; then $R(p(a_{1-e}))$ can be made arbitrarily small.

(20) Form $(q_i(a_{fi}))^t$ $(t = 1,2, \ldots)$ in $\mathfrak{R}(f_i)$, then

$$\lim_{t\to\infty} R((q_i(a_{fi}))^t) = 0.$$

A more qualitative interpretation of (13)–(20): Decompose a into $a_e + a_{1-e}$ or $a_e + \sum_i a_{fi}$. Then a_e is the *purely transcendental* part of a: In $\mathfrak{R}(e)$ every $p(a_e)$ is in the extreme case (α) (cf. above). a_{1-e} is the arbitrarily nearly algebraical part of a. Breaking it up further into the a_{fi}, we see even which irreducible polynomials express the algebraicity of a best. (13)–(20) can be used to build up a theory of proper values and of elementary divisors in $\mathfrak{R}$.

7. We obtained in §6 a complete answer to this question: How nearly 0 can $p(a)$ $(p(x)\epsilon P)$ be? That is: How small can $R(p(a))$ become? A different problem is this: How closely can a be approximated by elements a'

which are algebraic (with respect to $\mathfrak{Z}$)? That is: How small can $R(a-a')$ become for such elements a', for which $p'(x)\epsilon P$ with $p'(a') = 0$ exists?

We can answer this question, and it is remarkable that the answer is more satisfactory for continuous $\mathfrak{R}$'s than for discrete ones. We can prove by a rather involved analysis:

(21) If $\mathfrak{R}$ is continuous, then the algebraic elements (with respect to $\mathfrak{Z}$) are everywhere dense in $\mathfrak{R}$ (in the rank metric). That is: Given $a\epsilon\mathfrak{R}$ and $\epsilon > 0$, an $a'\epsilon\mathfrak{R}$ and a $p'(x)\epsilon P$ with

$$R(a - a') \leqq \epsilon,\ p'(a') = 0$$

exist. For a discrete $\mathfrak{R}$ (21) need not hold. In fact, if $\mathfrak{R} = \mathfrak{d}_n$, $\mathfrak{Z}$ = Center of $\mathfrak{R}$ = Center of $\mathfrak{d}$, then (21) is certainly false, if $\mathfrak{d}$ is not algebraic over its own center, $\mathfrak{Z}$![12]

Let us observe, finally, that every continuous $\mathfrak{R}$ contains "purely transcendental" a's, that is, a's with $\alpha = 1$, $\mathfrak{a} = \mathfrak{a}' = \mathfrak{R}$ (cf. (14)). For such an a $(p(a))^{-1}$ exists for all $p(x)\epsilon P$.

Integers. 8. Assume that a notion of *integers* is defined in $\mathfrak{Z}$. That is:

(I) Let a subring $\mathfrak{Y}$ of $\mathfrak{Z}$ be given, of which $\mathfrak{Z}$ is the quotient algebra. Use this terminology: The $a\epsilon\mathfrak{Z}$ are *rational*, the $a\epsilon\mathfrak{Y}$ are *rational integers*, the $a\epsilon\mathfrak{Y}$ with $a^{-1}\epsilon\mathfrak{Y}$ are *units*. An $a\epsilon\mathfrak{Y}$ which is no unit, but such that for $a = bc$, $b, c\epsilon\mathfrak{Y}$ either b or c is a unit, is a *prime*. a, b are *equivalent*, if ab^{-1} is a unit.

(II) Assume that every $a\epsilon\mathfrak{Y}$ can be written as

$$a = b_1 \ldots b_l, \quad b_1 \ldots b_l \text{ primes},$$

and that this representation is essentially (that is, up to permutations of the $b_1 \ldots b_l$, and replacement by equivalent ones) unique. We then define:

(22) A polynomial $p(x)\epsilon P$ that is

$$p(x) = x^l + \alpha_1 x^{l-1} + \ldots + \alpha_l,\ \alpha_1, \ldots, \alpha_l\epsilon\mathfrak{Z}$$

is *integer*, if the $\alpha_1, \ldots, \alpha_l$ are rational integers. An $a\epsilon\mathfrak{R}$ is an *algebraic integer*, if it is a (rank metric) limit point of algebraic integers. Observe how unreasonable this last definition would be, if we were dealing with real (or complex) numbers, and their usual (absolute value) metric: Every number would be a general integer in this sense. Here, however (that is in $\mathfrak{R}$ with its rank metric), we obtain a definite arithmetic structure, with properties which are obtained only after a detailed investigation.

We can prove:

(23) An $a\epsilon\mathfrak{R}$ is a general integer, if and only if all elements $q_1(x)$, $q_2(x)$, ... of $\mathfrak{J}$ in (12) are integer polynomials.

So the purely transcendental part of a (a_e in (18)) has no influence on a's arithmetical character.

From (23) we can derive rather directly:

(24) A general integer is algebraic, if and only if it is an algebraic integer.[13]

(25) A general integer is rational, if and only if it is a rational integer.[13]

We will develop the arithmetic based on these notions in subsequent publications.

[1] "Continuous Geometries," *Proc. Nat. Acad. Sci.*, **22**, 92–100 (1936); "Examples of Continuous Geometries," *Ibid.*, **22**, 101–108 (1936); "On Regular Rings," *Ibid.*, **22**, 707–713 (1936); "Algebraic Theory of Continuous Geometries," *Ibid.*, **23**, 16–22 (1937). These notes will be referred to as C. G.; E. C. G.; R. R.; A. T., respectively.

[2] Detailed accounts of the contents of the four notes quoted above,[1] with full proofs, were given by the author in lectures given in Princeton, in the years 1935–36 and 1936–37. Complete notes of these lectures were mimeographed. The entire theory will be presented as an Amer. Math. Soc. Colloquium Lecture in September, 1937, and also in book form in the Amer. Math. Soc. Colloquium Series.

[3] Cf. A. T., 20–21. "Regularity" is defined in R. R., 708, the lattice $R_{\mathfrak{R}}$ in R. R., 709–710, and A. T., 19.

[4] They must possess an "order" $\geqq 4$, cf. A. T., 20. This condition is necessary, in order to exclude "non-Desarguesian" (discrete, projective) plane geometries, which have only the order 3. Cf. loc. cit. above.

[5] For the notations cf. R. R., 707 (Def. 1) and 709 (Def. 5). As to the use of D and D', observe that $(a)_r$, $(a)_l{}^r \epsilon R_{\mathfrak{R}} = L$, and $(a)_l$, $(a)_r{}^l \epsilon L_{\mathfrak{R}} = L'$.

[6] In this case $\mathfrak{R}$ is the nth order matrix algebra over a suitable (not necessarily commutative but associative) division algebra $\mathfrak{d}$: $\mathfrak{R} = \mathfrak{d}_n$ (A. T., 17–18). Since $a \epsilon \mathfrak{R}$ is an nth order matrix over $\mathfrak{d}$, it possesses a matrix rank in the usual sense: $r(a) = 0,1, \ldots, n$ This is connected with the common geometrical dimension $d(\mathfrak{a}) = -1,0,1, \ldots, n-1$ of the linear subspace $\mathfrak{a}$ in L ($= L_n = R_{\mathfrak{R}}$) by the relation $r(a) = d((a)_r) + 1$. These $d(\mathfrak{a})$, $r(a)$ give our $D(\mathfrak{a})$, $R(a)$ by $D(\mathfrak{a}) = \dfrac{d(\mathfrak{a}) + 1}{n}$ (cf. C. G., footnote[13]) and $R(a) = \dfrac{r(a)}{n}$. So our rank $R(a)$ differs from the usual one $r(a)$ merely by the normalizing factor $\dfrac{1}{n}$.

[7] $R(a)$ is not an "absolute value" ("Bewertung") in the sense in which this notion is commonly used in algebra: For this the property (3), (ϵ) is too weak. For an "absolute value"

$$(\epsilon^*) \quad R(ab) \leqq R(a)\, R(b)$$

would be necessary. (3), (η), might do, although in most cases (e.g., in "p-adic" systems) even

$$(\eta^*) \quad R(a+b) \leqq Max + (R(a), R(b))$$

holds. Cf. J. Kurschak, *Jour. f. Math.*, **142**, 211–253 (1913); A. Ostrowski, *Acta Math.*, **41**, 271–284 (1918), and for the modern literature on this subject, e.g., W. Krull, *Jahresb. d. D. M. V.*, **46**, 153–171 (1936).

[8] J. H. Maclaggan Wedderburn, *Proc. London Math. Soc.*, **6**, 77–118 (1908), particularly 81, 98. See also v. d. W. II (cf. R. R., footnote[1]), 170.

[9] That is: Empty, or finite, or enumerably infinite.

[10] This number α ($\geqq 0$, $\leqq 1$) gives a quantitative estimate of how "nearly algebraic" a is, that is, how nearly $p(a) = 0$, $R(p(a)) = 0$, can be approximated.

[11] If the sum Σ is infinite at all (cf.[9]), then it converges with respect to the rank metric.

[12] In this case (21) with an $\epsilon \leqq 1/n$ requires $a = a'$, that is $p'(a) = 0$, that is the algebraicity (with respect to $\mathfrak{Z}$) of a itself.

[13] These are justifications of our use of the words "general integer." Note that these results would not hold if we considered real (or complex) numbers in their usual metric (cf. above, following (22)).